

# Objects, meaning, and the brain : neuroimaging studies on the impact of semantics on cortical integrative processing

Citation for published version (APA):

Döhrmann, O. (2009). *Objects, meaning, and the brain : neuroimaging studies on the impact of semantics on cortical integrative processing*. [Doctoral Thesis, Maastricht University]. Universiteit Maastricht. <https://doi.org/10.26481/dis.20090529od>

## Document status and date:

Published: 01/01/2009

## DOI:

[10.26481/dis.20090529od](https://doi.org/10.26481/dis.20090529od)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

# **Objects, meaning, and the brain**

Neuroimaging studies on the impact  
of semantics on cortical integrative processing

**Oliver Döhrmann**

**Colophon**

Layout design by Sarah Weigelt and Ellen Jongen

Cover design by Sarah Weigelt

ISBN 978 90 5278 832 6

Copyright © Oliver Döhrmann, Frankfurt am Main 2009

# **Objects, meaning, and the brain**

Neuroimaging studies on the impact  
of semantics on cortical integrative processing

## **Proefschrift**

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,  
op gezag van de Rector Magnificus, Prof. mr. G.P.M.F. Mols,  
volgens het besluit van het College van Decanen,  
in het openbaar te verdedigen  
op vrijdag 29 mei 2009 om 10.00 uur

door

**Oliver Döhrmann**

**Promotor**

Prof. dr. R. Goebel

**Copromotores**

Prof. dr. J. Kaiser

(University of Frankfurt, Germany)

Dr. M.J. Naumer

(University of Frankfurt, Germany)

**Beoordelingscommissie**

Prof. dr. B. Jansma (voorzitter)

Dr. E. Formisano

Prof. dr. T. Gruber

(University of Osnabrück, Germany)

Prof. dr. B. Roeder

(University of Hamburg, Germany)

The research reported in the present dissertation was conducted at the Institute of Medical Psychology, Goethe-University, Frankfurt am Main, Germany.

A Ph.D. scholarship by the German National Academic Foundation to Oliver Döhrmann is gratefully acknowledged.

# Contents

Chapter 1	Introduction	7
Chapter 2	Selectivity for animal vocalizations in the human auditory cortex	17
Chapter 3	Probing category selectivity to environmental sounds in the human auditory cortex	33
Chapter 4	Learning-induced cortical plasticity of audio-visual object representations	55
Chapter 5	Multisensory integration investigated by fMRI adaptation	77
Chapter 6	Semantics and the multisensory brain: How meaning modulates processes of multisensory integration	99
Summary		121
Samenvatting		125
References		129
Abbreviations		147
Acknowledgements		149
Curriculum Vitae		151

*By sight I have the ideas of light and colours, with their several degrees and variations. By touch I perceive hard and soft, heat and cold, motion and resistance, and of all these more and less either as to quantity or degree. Smelling furnishes me with odours; the palate with tastes; and hearing conveys sounds to the mind in all their variety of tone and composition. And as several of these are observed to accompany each other, they come to be marked by one name, and so to be reputed as one thing.*

George Berkeley (1710)

# 1

## Introduction



## The general idea

If people are asked to describe their immediate environment, it is very likely that they use words referring to visually perceived attributes of objects. For instance, they might start their description by providing names of these objects such as “car”, “house” or “street” and how these are generally arranged in a given setting. Maybe they will then go on to tell you more about specific features of some of these objects like their respective colors and shapes, if they are static or in motion and so on. However, it would be very surprising if someone replied to this rather open question by giving a description of which sounds are particularly dominant in a given scene – at least when this person’s sensory systems work in an undisturbed and healthy manner – and then proceeds to give a fine-grained analysis of their respective features such as the pitch or the perceived loudness. For most of us it seems to be more natural to rely especially on visual input to solve tasks of this kind. This is also reflected in the architecture of the human cortex in which at least twenty percent of its regions are sensitive to visual stimulation (Wandell et al., 2007).

Possibly, this visual dominance most of us experience every day is one explanation for the fact that research on perception and its neuronal underpinnings has similarly focused mainly on the visual system. However, even for processes related to object recognition the visual modality is not the exclusive source of information for a subject. More precisely, under conditions in which vision is deteriorated by darkness or dense fog one might still be able to perceive and even to recognize environmental objects, for instance by means of the sounds they emit. Furthermore, due to its impressing capabilities in the temporal domain the auditory system serves subjects very well for tasks in which objects have to be located in space (Scott, 2005; Thompson et al., 2006). In this dissertation, the focus will be particularly on the first of these aspects, namely those processes which are related to object recognition. The first two original studies collected in this dissertation (chapters 2 and 3) focused exclusively on the perception of environmental sounds and the respective cortical structures in the human brain involved in this task. I will elaborate more on this work in the next section of this introduction.

While it is an important and ambitious research project to obtain a complete picture of sensory perception and its cortical correlates for each sensory system in isolation, the overall aim of understanding how we actually experience our environment could possibly not be fully answered by such an approach. More precisely, in most cases our interactions with an object are based on unified experiences of a sensory “whole” which is constructed in the brain from different perceptual “streams” or sensory modalities. Thus, the default mode of perception in natural settings seems to be characterized by integration across our senses.

Back to the thought experiment concerning the description of an environment, it is quite intuitive that sensory modalities can either refer to the same aspect or can provide complementary information. An example of the former has already been given with regard to object recognition (an image and sound referring to the same object), and a second one might be based on features related to spatial coordinates, for instance motion processing which is possible both in vision and audition. In contrast, color and pitch are two distinct features of objects which could be regarded as com-

plementary, since these can only be derived from processing in the visual and the auditory sensory modality, respectively. If what is conveyed by different “channels” could, in principle, provide mutually supportive information, it is very likely that our brains were evolutionary adapted to exploit this fact for the construction of richer representations of the environment and objects embedded into it. A cortical representation of an object, like a dog or a hammer, is then built from multiple instances of contact with its sight and sound.

In the third section of this introduction I will provide more background information about two original empirical studies (chapters 4 and 5) which focused particularly on the question how processes of integration across the auditory and visual sensory modality are influenced by the respective semantic content of simultaneously presented sounds and pictures. Results from these two studies will then be synthesized into a first theoretical framework presented in the final chapter of this thesis. A brief outlook on this framework will be given at the end of this introduction.

The overall aim of this brief introductory chapter is to familiarize the reader with the overall aims of the present thesis and to provide a first conceptual framework. Therefore, to avoid unnecessary repetitions, more specialized information about the studies, for instance pertaining to methods and results as well as possible interpretations will be provided in the respective chapters. Generally, the four different empirical studies and the paper presenting the theoretical framework were all concerned with the question how the semantic content influences cortical processing of auditory (chapters 2 and 3) and audio-visual (chapters 4-6) object stimuli. In the following, “object stimuli” will be regarded as those sensory stimuli which originate from an environmental object. Furthermore, “meaningful object stimuli” can be operationally defined as those for which a (cortical) representation exists within a possibly widespread network of knowledge pertaining to the respective object (see e.g. Martin, 2007). We will later return to this notion.

## **Processing of auditory object stimuli**

Our work reported in chapter 2 investigated cortical correlates of the processing of animal vocalizations. Stimuli of this kind have several interesting features which enable the generation of an attractive stimulus set. For instance, they are typically very familiar to most of us, they could exert some behavioral relevance which increases aspects of ecological validity and they have a sufficient degree of complexity to tap into processes related to higher-level auditory processing. From the perspective of the present thesis particularly this last aspect is highly relevant. Incoming sounds are processed along a hierarchy starting on the level of subcortical nuclei and the brainstem and then involving various fields of the primary, secondary and higher-order auditory cortices (see e.g. Kaas and Hackett, 2000; Petkov et al., 2006). Similar to other sensory cortices, neurons at the earliest cortical processing stages are particularly tuned to relatively simple features. This is reflected in a corresponding topographical structure of these regions (“tonotopy”) which is also present in human subjects as revealed both by invasive (e.g. Howard et al., 1996) and non-invasive (e.g. Formisano et al., 2003) neuroimaging techniques. Moving “up” within the auditory “what” processing

stream demonstrates nicely that neurons show increasingly broader tuning properties with optimal responses, for instance, to broad-band noise stimuli (Rauschecker, 1998; Wessinger et al., 2001). These tuning profiles of neurons in higher-order cortices could be conceptualized as a kind of integration within the auditory sensory modality since representations of more complex stimulus features are generated from neuronal patterns related to simple stimulus properties. Finally, in the course of auditory object recognition or speech perception response patterns in higher-order sensory cortices are matched with “templates” or learned representations. It is this high level of processing, in which aspects of meaning for auditory object stimuli ultimately come into play. However, stimulus processing should not be regarded as a simple “feedforward sweep” along the subcortical and cortical hierarchy, but as a highly dynamic procedure involving a multitude of feedback loops and parallel processing. Thus, again, these dynamic interactions within the processing hierarchy could only be fully examined by using complex and, if possible, meaningful stimuli.

All these different aspects contribute to the more general notion that animal vocalizations form a category of stimuli which can be characterized by complex physical features and rich semantic content as derived from each subject’s learning history. In our first study, we aimed at examining if cortical regions particularly of higher-order auditory cortex selectively code for animal vocalizations by using an extension of the functional magnetic resonance imaging (fMRI) technique called fMRI adaptation (Grill-Spector et al., 2006). An experiment of this type investigates the impact of the first stimulus presentation on the processing of subsequently presented stimuli. Selectivity can be determined in an fMRI-adaptation paradigm by comparing if the measured signal in a cortical region differentiates between experimental conditions involving either stimulus repetitions or changes. A stronger signal for the latter compared to former conditions can be regarded as an index of sensitivity for those features of a stimulus which were varied. In our study reported in chapter 2 we asked which cortical regions exhibited fMRI-adaptation effects as determined by signal differences between repeated and different animal vocalizations. Importantly, in sensory cortices adaptation could also be induced by the repetition of features common to many auditory stimuli. This underscores the need to construct appropriate control conditions to strengthen accounts claiming stimulus-selective representations in a cortical region. An approach to do so is described in more detail in chapter 2. However, it should be further noted that findings and even more so null effects in fMRI-adaptation experiments are in many respects heavily dependent on the experimental design. Thus, depending on the neuronal tuning properties in a particular region of interest the choice of the appropriate type of paradigm could critically determine the successful search for fMRI-adaptation effects (for a review on these issues see Weigelt et al., 2008).

In a second auditory study (chapter 3) we went on to examine if those cortical regions found to be sensitive to animal vocalizations in terms of an fMRI-adaptation effect were modulated differently by the presentation of sounds referring to inanimate objects (tools). More precisely, if those cortical regions which showed an adaptation effect to animal vocalizations in the first experiment did not adapt to repeated tool sounds, this would further enhance the notion of selectivity to this particular stimulus category. The general rationale for this study design is derived from the finding that some higher-order cortical regions involved in sensory processing show particu-

larly strong responses to certain stimulus categories. The so-called fusiform face area (Kanwisher et al., 1997) is probably the most well-known cortical region of this kind. However, not only visual stimuli, but also different auditory categories such as animal and tool sounds have been associated with distinct cortical foci, in this case located in the lateral temporal lobe (Lewis et al., 2004; 2005). Although it might be less likely that these auditory foci form special “modules” dedicated solely to the processing of one particular sound category, this category-related functional differentiation is still a highly interesting research area. Accordingly, besides the aim of further investigating the question of selectivity of some cortical regions for animal vocalizations, we were generally able to examine if animal and tool sounds produced corresponding adaptation effects in those cortical regions identified with conventional fMRI experiments.

The two fMRI-adaptation studies reported in this dissertation have to be regarded as part of a bigger research project aiming at a better grasp of auditory object recognition. Although both involved semantically “loaded” stimuli and in the second study additionally stimuli from two different categories, they necessarily can only provide small pieces of information with respect to this research project. Moreover, since they operated on an “intermediate” level at the intersection of perception and cognition, further studies will have to shed more light both onto the contribution of lower-level stimulus features and of higher-order conceptual processes to the cortical response patterns detected in our studies.

## **Audio-visual object processing**

Classical experimental work on questions related to the integration across sensory modalities (crossmodal integration) focused on so-called multisensory neurons, which are generally responsive to unimodal auditory and visual as well as to bimodal audio-visual stimuli (Stein and Meredith, 1993; Stein and Stanford, 2008). These early studies demonstrated the influence of spatial and temporal factors during integrative processing of multisensory neurons as measured by increases in their spiking rate for bimodal stimulation matching in time and space compared to each response and even the sum of responses to unimodal stimulation. Effects of this kind can be explained very well by the receptive field properties of multisensory neurons. More precisely, temporally coinciding and spatially corresponding audio-visual stimuli induce stronger spiking activity in these neurons than stimuli which are spatially misaligned. The latter situation could even suppress the activity patterns of these neurons below the level of responses to unimodal stimulation due to inhibitory mechanisms (Stein and Stanford, 2008).

Results like that were particularly obtained by research utilizing animal models such as the cat. Highly interesting data were additionally generated by shifting the perspective from a fully developed organism to the situation of neonatal subjects. For instance, rearing cats without any visual experience had no consequences for the spiking activity of multisensory neurons when unimodal visual or auditory stimuli were presented. However, the situation was dramatically altered in the case of bimodal stimuli: the multisensory neurons of these cats did not show the typical increase in spiking activity which would have indicated sensory integration (Wallace et al., 2004).

Thus, although the neonatal brain seems to be equipped with potentially integrative multisensory neurons, the respective processing critically depends on an interaction with a typical sensory environment.

In human subjects, developmental work suggests that matching temporal and spatial information conveyed by both the auditory and the visual sensory modalities could guide the learning even of arbitrary associations between the sight and the sound of particular objects. For instance, Gogate and Bahrick (1998) demonstrated that infants at the age of 7 months were capable of learning associations between objects and speech sounds when both were presented in temporal synchrony, but not for static objects or those moving asynchronously to the vowels. Furthermore, changes of the temporal synchrony, from an initial study phase to an experimental phase could already be detected by infants at 3.5 months (Bahrick, 1992) or possibly even by newborns (Morrongiello et al., 1998). In contrast, changes in arbitrary relations between the color or form of an object and high- or low-pitched sounds seemed to be perceived only by about 7 month-olds (Bahrick, 1994). Together, these findings suggest that at a very early developmental level human subjects are already able to use spatial and temporal cues for audio-visual integration. However, only after at least half a year of experience with the respective sensory environment infants can successfully start to exploit information related to the content of stimuli for crossmodal integrative processing.

Very often and most likely in the course of a child's development, information about sensory features of objects is further enhanced by many other pieces of information, for instance pertaining to the proper use (for tools) or to cognitive operations such as categorization ("dog" vs. "fox"). All these different aspects generate potentially large networks of explicit and implicit "knowledge" with their respective cortical correlates (Martin, 2007; Patterson et al., 2007). While different sensory modalities provide important means for generating these cortical representations ("bottom-up"), there is also some influence of the latter on perceptual processing ("top-down"). For instance, if one considers again a situation of poor viewing conditions in which an object is barely visible, a (simultaneous) matching sound could greatly improve object recognition. Similar results such as a behavioral gain in terms of reaction times for "semantically" congruent audio-visual stimuli were revealed also under optimal perceptual conditions (e.g. Laurienti et al., 2004).

With this general framework in mind, it is now possible to outline the second part of the present dissertation. In this part I will report work from two empirical studies we conducted to investigate audio-visual processing of environmental stimuli. Importantly, the aim of these studies was to contribute to a better understanding of how the formation and later on the presence of cortical object representations can exert influences on the processing of subsequently presented audio-visual stimuli. The first audio-visual fMRI study (chapter 4) had the general aim to simulate the situation which is very common for many children and even for adults whenever the sight and sound of a previously unknown object is experienced for the first time and then more often during subsequent encounters. Subjects were scanned twice, before and after the training of associations between previously unrelated artificial auditory and visual stimuli. We were interested in examining which cortical regions showed plastic changes related to the generation of novel audio-visual associations. Apart from this learning-induced

manipulation we additionally varied the degree of congruency between the image and the sound both for familiar natural (animal) as well as for artificial stimuli. For animal stimuli, cortical responses to semantically congruent and incongruent audio-visual stimulation could be compared in both fMRI sessions of that study. In contrast, only after introducing at least a basic degree of congruency by associative learning we could examine similar issues for the artificial stimuli. Accordingly, we asked if the cortical response patterns for these two radically different types of stimuli converged to some extent after learning.

Our first study showed that factors such as familiarity and meaning of stimuli are important for object-related audio-visual processing (see also Hein et al., 2007). The data presented in chapter 4 added to the existing evidence (e.g. Ernst and Bühlhoff, 2004) that the brain is “tuned” to exploit statistical regularities to solve binding problems even across sensory modalities. Effects measured during that study were induced after an explicit instruction to learn associations between the sight and the sound of a novel artificial object. Other types of processing may operate on a short-term level and even unintentionally (see Shams and Seitz, 2008 for a recent review on crossmodal learning). Results from the first audio-visual study suggested that particularly deviations from previously learned bimodal associations (i.e. long-term memory representations) induced a response pattern in cortical regions commonly reported for conflict processing.

In a next experimental step (chapter 5) we were interested to learn which cortical regions are sensitive to short-term changes between two successive audio-visual stimuli. To this end, we again employed an fMRI-adaptation paradigm presenting successive audio-visual object stimuli varying either the auditory, the visual, both or none of the components from the first to the second presentation. In contrast to the previous experiment, we did not incorporate an additional variation of semantic congruency although we again used a stimulus set consisting of images and sounds of animals. Thus, changes did not involve the additional occurrence of incongruency, neither within the second stimulus nor across the two audio-visual stimuli.

This study forms the basis for future work which will again introduce manipulations of semantic congruency for object stimuli, but then in the course of fMRI adaptation paradigms. Only a few previous studies have employed the fMRI adaptation technique in the crossmodal domain, and we feel that its potential to reveal complementary results compared to conventional designs has not been fully exploited yet. Indeed, the study reported in chapter 5 is to our knowledge the first to examine repetition-related effects for audio-visual object stimuli in the course of an fMRI-adaptation paradigm.

Although in the last couple of years some groups have started to work on issues related to the question of the exact role of semantics for crossmodal integration, methodological advances such as the use of fMRI adaptation will certainly further refine and possibly also redefine existing theoretical accounts. The final chapter of this thesis (chapter 6) offers a first conceptual framework for the systematization of the current state of research on the impact of semantics on audio-visual integration. Based on a review of studies which involved variations of the semantic content in the course of crossmodal experiments, we suggest that processing of semantically congruent and incongruent audio-visual stimuli is accomplished by two distinct networks of cortical regions. As mentioned above, semantic incongruency most likely taps into processes of

conflict processing. The latter is associated with activation of fronto-parietal regions. In contrast, semantically matching audio-visual stimulation seems to generate responses which could be found predominantly in more posterior and more ventral parts of the brain, very often in close proximity to sensory cortices.

We hold that this theoretical framework could provide a first guide for the generation of hypotheses for future experiments. However, it is necessarily far from being complete. Research both in the unimodal and the crossmodal domains face novel challenges which have to be answered by an extended set of neuroimaging methods. Particularly crossmodal studies will have to provide answers to the question which indices convincingly underscore that sensory integration is indeed taking place in a given cortical region. Due to technical limitations we presently cannot search non-invasively for multisensory neurons in the human cortex to strengthen accounts derived from neuroimaging experiments. Approaches which compared fMRI responses induced by unimodal and bimodal stimulation have generated a large body of literature, but also a discussion about the appropriateness of certain methodologies for the definition of cortical integration regions (see e.g. Calvert, 2001 and Beauchamp, 2005 for two possible standpoints). It will be exciting to see how novel experimental and analytical methods can shed further light onto these issues.

Finally, research has only recently started to delve into a question which is deeply connected to the previous issue: Novel techniques can only provide answers when researchers are able to ask precise questions. Thus, using advanced experimental paradigms will not transform our present knowledge of the multisensory brain when this development is not accompanied by the generation of novel theoretical frameworks and innovative research questions. In the last couple of years the existing view of cortical organization with a strict division into unisensory, multisensory (associative) and non-sensory regions (e.g. Mesulam, 1998) has been called into question by findings suggesting at least interactions between sensory modalities on the earliest cortical processing levels (Schroeder and Foxe, 2005; Kayser and Logothetis, 2007). Some researchers (Ghazanfar and Schroeder, 2006) have even suggested the bold hypothesis that the complete neocortex is “essentially multisensory”. Our work in the audio-visual domain at least supports the view that processing in different sensory modalities is more tightly coupled than previously regarded. However, to what extent multisensory processing is indeed taking place at different processing levels in the human brain will certainly be an important line of future research in which intriguing new experimental techniques will be paired with profound conceptual advances (see also Driver and Noesselt, 2008).

## Summary and concluding remarks

The overall aim of the present thesis is to shed more light on the impact of semantics on integration-related processing within the auditory sensory modality and across the auditory and visual processing streams. Since the influence of stimulus content is particularly pronounced in the course of object recognition, the studies collected in the present thesis consequently found cortical regions in the temporal and frontal lobes implicated in the so-called “what-streams”. However, it is well-known that “semantics” is



likely to involve further aspects as could be manipulated in the course of the present thesis (Martin, 2007). For instance, it has been repeatedly shown that for stimuli associated with particular actions (such as tools) cortical regions related to the respective object use become co-activated with those involved in recognition (Lewis, 2006). This demonstrates that we have only begun to understand how the brain is able to extract meaning from stimuli and how this, in turn, influences subsequent stimulus processing.

Ultimately the following chapters are inspired by the major question of how the dynamics of neuronal activity found for the processing of complex stimuli becomes integrated in what we experience as a “perceptual whole” (Engel et al., 1999; Singer, 1999). An answer to this intriguing question will require joint forces by multimethod approaches for the generation of new data and novel theoretical frameworks for their interpretation.





# 2

## Selectivity for animal vocalizations in the human auditory cortex

We aimed at testing the cortical representation of complex natural sounds within auditory cortex using human functional magnetic resonance imaging (fMRI). To this end, we employed two different paradigms in the same subjects: a block-design experiment was to provide a localization of areas involved in the processing of animal vocalizations, while an event-related fMRI adaptation experiment was to characterize the representation of animal vocalizations in the auditory cortex. During the first experiment, we presented subjects with recognizable and degraded animal vocalizations. We observed significantly stronger fMRI responses for animal vocalizations compared to the degraded stimuli along the bilateral superior temporal gyrus. In the second experiment, we employed an event-related fMRI-adaptation paradigm in which pairs of auditory stimuli were presented in four different conditions: a) two identical animal vocalizations, b) two different animal vocalizations, c) an animal vocalization and its degraded control and d) an animal vocalization and a degraded control of a different sound. We observed significant fMRI adaptation effects within the left STG. Our data thus suggest that complex sounds such as animal vocalizations are represented in putatively non-primary auditory cortex in the left STG. Their representation is probably based on their spectro-temporal dynamics rather than simple spectral features.

Altmann, C. F., Doehrmann, O., and Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex*, 17, 2601-2608.

## Introduction

The ability to quickly identify complex sounds enables us to interact adequately with our environment when visual information is scarce. Neuropsychological studies have described brain lesions that lead to impaired processing of complex auditory sounds (Engelien et al., 1995; Clarke et al., 2000, 2002). This deficit - auditory agnosia - has been associated with left, right or bilateral lesions of temporal or temporo-parietal cortex, suggesting specialized cortical networks for complex sound processing. However, the coarse localization of lesions has prevented the identification of the exact anatomical correlates of higher-order auditory processing. Neurophysiological studies in non-human primates provided evidence for hierarchical processing in macaque primary and non-primary auditory cortex (Rauschecker et al., 1995; Kaas et al., 1999; Rauschecker and Tian, 2000; Tian et al., 2001). More specifically, histological studies described a sub-parcellation of macaque auditory cortex into core and adjacent belt regions (Morel et al., 1993; Jones et al., 1995; Hackett et al., 1998). Functionally, neurons in the core areas have been shown to respond well to pure tones (Rauschecker et al., 1997), while belt areas show a preference for more complex stimuli such as band-passed noise (Rauschecker et al., 1995). Moreover, neural populations in lateral belt areas appear to be selective for both the rate and the direction of frequency-modulation (Rauschecker, 1997, 1998a, 1998b) and for monkey vocalizations (Rauschecker and Tian, 2000). For human subjects, a similar hierarchy of auditory processing has been reported, but the actual homologies are still under debate. In particular, anatomical and histochemical studies have described a subdivision into core, belt and parabelt auditory cortex in the superior temporal lobe, homologous to the macaque anatomy (Sweet et al., 2005). Recent fMRI studies have revealed the existence of at least 2 tonotopically organized areas on the superior temporal plane, putatively corresponding to macaque A1 and R (Formisano et al., 2003). Furthermore, a central core area showing tonotopic organization is surrounded by belt areas that exhibited stronger fMRI responses to band-pass noise compared to pure tones (Wessinger et al., 2001; Seifritz et al., 2006). For non-primary auditory cortex lateral and posterior to Heschl's gyrus, recent fMRI studies have observed stronger responses to amplitude and frequency-modulated tones compared to unmodulated tones (Hall et al., 2002; Hart et al., 2003). Interestingly, the areas that exhibit the two types of modulation are largely overlapping, suggesting common neural correlates. Both amplitude and frequency modulation are properties of complex environmental sounds that might serve as basic features to process these sounds. A series of fMRI studies presented subjects with complex natural sounds (animal voices, tools, dropped objects and liquids) and compared the fMRI responses for recognizable sounds with unrecognizable reversed sounds (Lewis et al., 2004). The findings of these studies suggested that a predominantly left-hemispheric semantic network is involved in the processing of recognizable complex sounds, in addition to bilateral foci in the posterior portions of the middle temporal gyrus. When different categories of natural sounds were compared, stronger fMRI activity was found in a predominantly left-lateralized cortical network including the left and right posterior portions of the middle temporal gyrus for tools versus animal sounds (Lewis et al., 2005). In contrast, the bilateral middle portion of the superior temporal gyrus (STG) exhibited a preference for animal compared to tool sounds.

Another important function of the human auditory system is the processing of speech sounds. For example, stronger fMRI responses to speech compared to frequency-modulated tones have been demonstrated in the STS (Binder et al., 2000). However, in the same study, similar fMRI responses were obtained for reversed speech and pseudo-words, suggesting that these areas are responsive to the acoustic properties of speech rather than language per se. Interestingly, results from an fMRI study that compared cortical responses to consonant-vowel syllables to pure tones and noise have suggested an implication of bilateral STS and the planum temporale in language-specific processes (Jäncke et al., 2002). In line with this evidence, by employing sine-wave-speech stimuli, a recent fMRI study has shown involvement of the left posterior STS in speech recognition (Möttönen et al., 2006).

The comparison of different categories is able to reveal brain regions with differential preferences for one category over the other. However, this approach may not be appropriate to identify areas that are involved in the processing of more than one category. In particular, an area that has spatially overlapping or close neural populations encoding different stimulus categories would not show differential fMRI responses due to the limited spatial resolution of fMRI and would thus be regarded as not involved in the encoding of the respective categories. To overcome this problem, the present study employed an fMRI adaptation paradigm, which takes advantage of the observation that fMRI responses decrease with repeated presentation of the same stimuli. This technique has been widely applied in the investigation of the human visual system (for a review: Grill-Spector et al., 2006). For example, a recent fMRI adaptation study (Altmann et al., 2004) has tested the effect of visual context on shape processing in higher visual areas. For the auditory domain, the evidence for fMRI adaptation effects is rare. For instance, adaptation effects for speaker's voice have been demonstrated in the anterior part of the right superior temporal sulcus (Belin and Zatorre, 2003). Furthermore, a recent study has described correlations between behavioral effects of repetition priming and repetition-associated reduction of fMRI responses to environmental sounds in the right superior temporal gyrus, the bilateral superior temporal sulci and right inferior frontal gyrus (Bergerbest et al., 2004). Previous studies that tested for repetition priming effects in the auditory domain were not able to reveal significant reductions in the fMRI signal (Buckner et al., 2000; Wheeler et al., 2000; Badgaiyan et al., 1999, 2001). Bergerbest et al. (2004) argued that this lack of repetition-related reductions could possibly be explained by the fact that these studies have used auditory word-priming tasks that rely more on phonological rather than acoustic representations. For this reason, it might be more adequate to employ environmental sounds rather than spoken words to test for repetition effects in the auditory cortex.

As the processing of different categories of auditory objects differs in its spatial (Lewis et al., 2005) and temporal (Murray et al., 2006) properties, we decided to test fMRI-adaptation effects for a single category of complex sounds only, i.e. animal vocalizations. This class of environmental sounds is usually experienced and learned early in child development and refers to tangible and perceptually rich objects in our environment. Thus, the present study used human fMRI to investigate the neural mechanisms that underlie the representation of animal vocalizations as a specific category of complex natural sounds. To this end, we employed two different techniques: first, we conducted a block-design experiment for a coarse localization of cortical areas that are

involved in the processing of animal vocalizations. Second, we employed an fMRI adaptation paradigm to test for the form of representation of animal vocalizations within the auditory cortex. In the block-design experiment, we presented subjects with animal vocalizations and control sounds with similar spectral features but distorted temporal structure. We manipulated the degree to which the animal vocalizations were degraded and tested for a correlation between the degree of scrambling with BOLD responses in the human auditory cortex. In the fMRI adaptation paradigm we tested for selective representation of animal vocalizations. To this end, we presented pairs of sounds that were a) two identical animal vocalizations, b) two different animal vocalizations, c) an animal vocalization and its degraded control and d) an animal vocalization and a degraded control of a different animal vocalization. Brain areas that selectively represent animal vocalizations should exhibit stronger fMRI responses for different compared to same animal vocalizations. In contrast, areas that mainly represent simple spectral features should exhibit cross-adaptation when an animal vocalization and its degraded control is presented and a rebound from adaptation in response to the combination of an animal vocalization and a degraded control of a different animal vocalization.

## Materials and Methods

### Subjects

Thirteen healthy, right-handed volunteers (age range 19-38, 8 male) participated in the two experiments. Handedness was evaluated by self-report. All subjects had normal hearing abilities and gave their informed consent to participate in the study. The experiments were performed in accordance with the ethical standards laid down in the 1964 declaration of Helsinki.

### Stimuli

Ten different animal vocalizations (cat, chicken, cow, dog, donkey, duck, frog, horse, owl, sheep) were taken from a database specifically designed for auditory psychophysics (Marcell et al., 2000). Sounds were digitized with a sampling rate of 44,100 Hz. Sound duration was on average 990 msec (standard deviation 180 msec, range: 650-1250 msec), sound intensity was individually adjusted to a comfortable level resulting in an average sound pressure level of 86 dB. The sounds were equalized as regards their root mean square energy and presented with MR-compatible head phones (Resonance Technology, Northridge, USA). The frequency response of the head phones ensured reliable presentation of the employed stimuli. More specifically, the frequency response was maximal at about 1.5 kHz (99 dBC), remaining at about 82-88 dBC for lower frequencies (0.25-1 kHz), and at about 83-88 dBC for higher frequencies (2-8 kHz).

We verified the recognizability of the animal vocalizations in a separate session, instructing 6 naive subjects to spontaneously name each animal. We employed the same auditory stimulus system as used in our fMRI experiments. The naming performance showed that the animals were reliably recognized (cat: 6/6; chicken: 6/6; cow: 6/6; dog: 6/6; donkey: 6/6; duck: 3/6; frog: 6/6; horse: 6/6; owl: 6/6; sheep: 6/6).

As shown in Figure 1, control stimuli consisted of degraded versions of the animal vocalizations that were generated in the following way: we performed a short-term Fourier transform on each stimulus (windows of 64 time points), and increasingly degraded the animal vocalizations by permuting the values of the amplitude and phase components across time (0%, 25%, 50%, 75% and 100% degradation). Then, we performed the inverse discrete Fourier transform on that signal to recover the amplitude waveform (Fig. 1). The resulting stimuli preserved the average output frequency and similar overall spectral composition. The temporal structure however, that is the amplitude envelope and periodicity was disrupted. Previous fMRI studies employed a similar approach, either permuting the amplitude and phase values across frequencies, preserving the overall temporal structure (Belin et al., 2002) or generating spectrally matched noise (Thierry et al., 2003) with disrupted temporal structure.

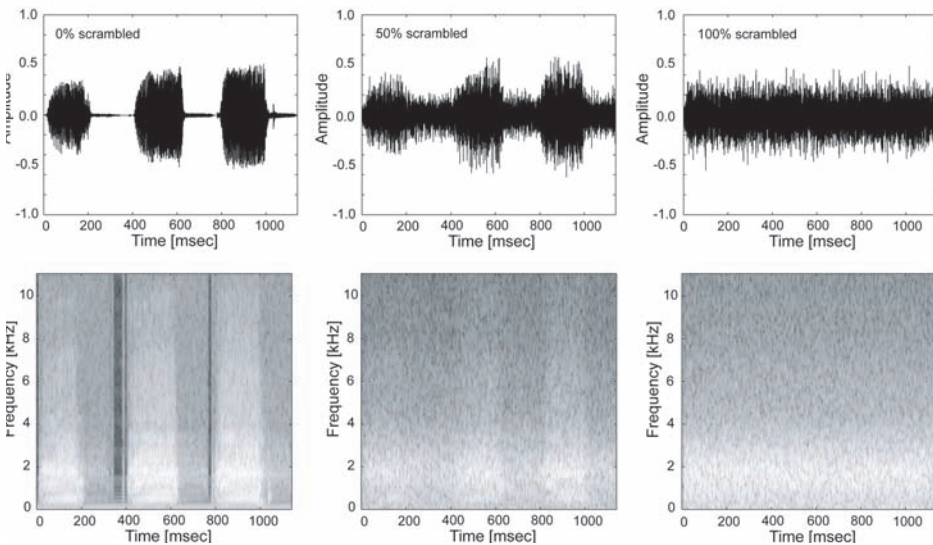


Figure 1. Sample Stimulus (“duck vocalization”). Amplitude waveform (upper panel) of the original stimulus (left), 50% degraded (middle) and 100% degraded (right). Spectrograms (lower panel) of the original stimulus (left), 50% degraded (middle) and 100% degraded (right).

## Procedure

All subjects were administered two runs of the block-design experiment, four runs of the event-related fMRI-adaptation experiment and a high-resolution T1-weighted MPRAGE sequence as an anatomical reference. The block-design experiment consisted of 15 blocks which lasted 30 seconds each with a 15 second stimulation period and 15 second scanner noise only period without additional auditory stimulation (baseline). Within each stimulation block, 10 sounds were presented with an inter-stimulus-interval of 1.5 seconds. Five different conditions, that is five different degrees of degradation (0%, 25%, 50%, 75% and 100% degraded) were presented block-wise. During the full run that lasted 7 minutes and 40 seconds, subjects were instructed to keep their

eyes open and fixate a white cross on a black background, back-projected by a LCD-projector through a mirror fixed to the head coil. Within each block, one single stimulus was repeated in sequence. Subjects were instructed to respond to this repetition with a button press to ensure attentive processing of the stimuli.

For the event-related fMRI experiment, an experimental run was comprised of 80 pairs of auditory stimuli. Each experimental run had a duration of 8 min 30 sec and contained four different conditions: a) two identical animal vocalizations, b) two different animal vocalizations, c) an animal vocalization and its degraded control (100% degraded) and d) an animal vocalization and a degraded control (100% degraded) of a different animal vocalization. Within a single run, 20 trials per condition were presented, intermixed with 20 trials without auditory stimulation (baseline). Similar to previous studies (Altmann et al., 2004), the order of presentation was counterbalanced so that trials from each condition were preceded (1 trial back) equally often by trials from each of the other conditions. The inter-trial-interval was 5 seconds, the delay between first and second stimulus was 510 msec on average. To ensure efficient parameter estimation of the hemodynamic response function, the onset of the first stimulus was jittered by  $\pm 500$  msec (Dale et al., 1999). During the whole experiment, subjects were instructed to keep their eyes open and look at a fixation cross. Subjects were required to press a button with one hand when a stimulus pair contained a degraded control stimulus and a different button with the other hand when a stimulus pair contained two animal vocalizations. Subject showed above 98 per cent correct responses in all conditions at this task.

## Imaging and Data Analysis

Imaging was conducted on a 3T Siemens ALLEGRA scanner at the Brain Imaging Center Frankfurt, Germany. An echo planar imaging sequence with gradient echo sampling (TR = 2 sec, TE = 35 msec) was used to acquire the functional imaging data. Thirty-four axial slices (3 mm thickness with 3.00 x 3.00 mm in-plane resolution, inter-slice gap: 0.3 mm), covering the whole brain, were collected with a head coil. The field of view was 19.2 cm x 19.2 cm with an in-plane resolution of 64x64 pixels. As anatomical reference, we acquired 3D volume scans by using a magnetization-prepared rapid-acquisition gradient echo sequence (160 slices; TR: 2300 ms).

fMRI data were pre-processed and analyzed using the Brainvoyager QX (Brain Innovation, Maastricht, Netherlands) software package. Preprocessing of functional data included slice scan time correction, head movement correction and linear detrending. The 2D functional images were co-registered with the 3D anatomical data, and both 2D and 3D data were spatially normalized into the Talairach and Tournoux stereotactic coordinate system (Talairach and Tournoux, 1988). Functional 2D data were spatially filtered employing a Gaussian filter with 4 mm FWHM for the group analysis only. The experimental effects in the different conditions were compared employing a random-effects general linear model. To display statistically significant differences, we employed a criterion of  $p < 0.001$ , uncorrected for multiple comparisons. This criterion was chosen to be equal to the constant criterion used in the single subject ROI selection for experiment 2. To achieve anatomical correspondence and to reduce the inter-subject variability of cortical gyrification, we applied cortex-based inter-subject alignment (Fischl et al., 1999).

For the event-related fMRI adaptation experiment, we selected three regions of interest (ROI) for each individual subject. An area within the STG was selected functionally on the basis of Experiment 1. More specifically, for each individual subject, we selected all voxels within the bilateral STG that showed a significant decrease of the fMRI responses ( $p < 0.001$  uncorrected) with the degradation of the animal vocalizations but significantly stronger responses for all sounds compared to the baseline stimulation in Experiment 1. Heschl's gyrus (HG) and the planum temporale (PT) were determined in a first step anatomically, employing the anatomical landmarks of the cortical surface of each individual subject (Brechtman et al., 2002). Within the anatomically defined areas, we selected functionally those areas that showed significantly stronger fMRI responses for all auditory stimuli of Experiment 1 compared to the baseline. This procedure aimed at avoiding averaging across neural populations that are tuned to frequencies beyond the range of the employed stimuli. Finally, our ROI analysis in HG and PT was restricted to those voxels that fell within the 90% boundaries of the probability map for HG (Rademacher et al., 2001) and the 95% boundaries for PT (Westbury et al., 1999), respectively. All ROIs were selected not to overlap with each other in order to avoid dependencies between them. For all ROIs, we conducted a general linear model to estimate the beta weights that model the hemodynamic responses as a combination of two gamma functions within these areas (Friston et al., 1995; 1998).

The beta weights were related to the overall signal level to express them as percent signal change values. In Figure 3, the standard error bars of the fMRI responses were corrected for inter-subject effects similar to previous studies (Altmann et al., 2004). In particular, the fMRI responses were calculated individually for each subject by subtracting the mean percent signal change for all the conditions within this subject from the mean percent signal change for each condition and adding the mean percent signal change for all the conditions across subjects. We then compared the percent signal change values of all subjects across conditions by a one-way ANOVA with repeated measurements. The degrees of freedom were corrected with the Greenhouse-Geisser method when appropriate. In case of a significant ANOVA result, we conducted further linear contrast analysis to compare fMRI responses to same and different animal vocalizations and to compare pairs with control stimuli generated from same or different animal vocalizations. This analysis aimed at characterizing the adaptation effects for animal vocalizations and to test for cross-adaptation from an animal vocalization to its control.

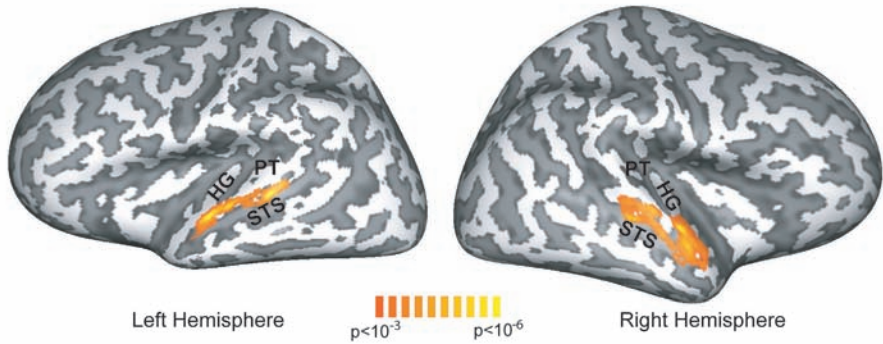
## Results

In Experiment 1, we tested the neural correlates of complex sound processing by investigating the relationship between the parametrically manipulated degradation of animal vocalizations and the acquired fMRI responses. We employed a random-effects general linear model to characterize common fMRI responses to animals versus degraded control stimuli across the subjects ( $n = 13$ ). As shown in Figure 2a, linear contrast analysis revealed bilateral areas in the superior temporal gyrus (STG) that exhibited a significant ( $p < 0.001$  uncorrected) decrease of fMRI responses with degradation of



animal vocalizations. As depicted on Figure 2b, these effects were robustly observed within individual subjects.

**A** Group data for Experiment 1



**B** Data for Experiment 1 of a single subject

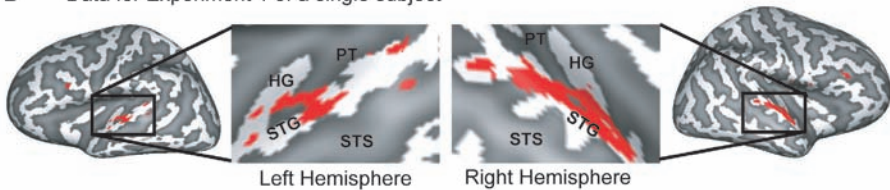


Figure 2. (A) Group data ( $n=13$ ) of the cortex-based general linear model analysis of Experiment 1 shown on an inflated 3D reconstruction of the cortical surface of a single subject which served as the target for cortex-based inter-subject alignment. Gyri are shown in light gray, sulci in dark gray. The activation map shows voxels that exhibited significantly ( $p<0.001$ , uncorrected) stronger fMRI responses for intact versus degraded stimuli as determined by linear contrast analysis (contrast weights: 0% degraded: +2; 25% degraded: +1; 50% degraded: 0; 75% degraded: -1; 100% degraded: -2). (B) Data for a representative subject for Experiment 1. This subject also served as target for the cortex-based alignment procedure employed in the group analysis. Voxels that exhibited significantly ( $p<0.001$ , uncorrected) stronger fMRI responses for intact versus degraded stimuli are depicted in red. HG: Heschl gyrus; PT: planum temporale; STG: superior temporal gyrus; STS: superior temporal sulcus.

To assess the selective representation of animal vocalizations versus control stimuli with similar spectral features, we conducted an event-related fMRI-adaptation experiment. We employed ROI analysis to characterize the fMRI response profile within the auditory cortex for this experiment. The ROIs were based on Experiment 1, ROI sizes and locations are shown in Table 1. Left and right HG did not differ significantly in size (paired two-tailed t-test:  $t[12] = 1.92$ ;  $p = 0.08$ ), similar to left and right STG ( $t[12] = 1.38$ ;  $p = 0.19$ ). However, the left PT ROI was significantly larger than right PT ( $t[12] = 3.78$ ;  $p<0.01$ ).

Table 1. Regions of interest

	Talairach			ROI-size (voxels)
	x	y	z	
<b>ROI</b>				
HG-right	45 +/- 4	-15 +/- 6	8 +/- 4	53
HG-left	-44 +/- 3	-20 +/- 4	8 +/- 2	62
PT-right	52 +/- 4	-28 +/- 5	14 +/- 4	34
PT-left	-50 +/- 3	-34 +/- 5	14 +/- 4	49
STG-right	57 +/- 3	-15 +/- 7	4 +/- 4	63
STG-left	-57 +/- 3	-20 +/- 5	5 +/- 3	50

Talairach coordinates in mm +/- standard deviation averaged across subjects for the selected Regions of Interest (ROIs) that had significantly stronger fMRI responses in Experiment 1 for auditory stimuli versus baseline (HG, PT) or for animal vocalizations compared to degraded control stimuli (STG).

As shown in Figure 3, within the areas in the STG that exhibited significantly stronger fMRI responses for animal vocalizations versus degraded sounds, we observed significant differences across experimental conditions for the adaptation experiment. In particular, a repeated measurements ANOVA on the percent signal change levels as shown in Figure 3b) with the factor Condition (Same Animal Vocalizations, Different Animal Vocalizations, Animal – Same Control, Animal – Different Control) revealed significantly different fMRI responses across conditions in the left STG ( $F[3,36] = 12.89$ ;  $p < 0.001$ ), but not for the right STG ( $F[3,36] = 2.59$ ;  $p = 0.07$ ). A linear contrast analysis in the left STG showed significantly stronger fMRI responses for different versus same animal vocalizations ( $F[1,12] = 9.70$ ;  $p < 0.01$ ), indicating a significant rebound from adaptation. The interaction between the factors Hemisphere (left/right STG) and Repetition (same/different animal vocalization) was significant ( $F[1,12] = 12.39$ ;  $p < 0.05$ ), suggesting differential processing across the two hemispheres.

We furthermore compared conditions with animal vocalizations only and pairs of animal vocalizations and control stimuli. A repeated measurements ANOVA revealed significantly lower fMRI responses for the pairs that contained control stimuli within bilateral STG (left STG:  $F[1,12] = 20.05$ ;  $p < 0.001$ ; right STG:  $F[1,12] = 6.90$ ;  $p < 0.05$ ). Since the ROIs were selected on the basis of the previous block-design experiment as areas with stronger responses for animal vocalizations versus control stimuli, it was expected that they would also show this effect in the fMRI adaptation experiment. However, if these ROIs were selective for simple spectral features rather than the temporal structure of stimuli, we should also expect fMRI adaptation effects for the conditions when the animal vocalizations are paired with their degraded versions compared with different control stimuli. We did not observe significant differences between same or different control stimuli in the STG (left hemisphere:  $F[1,12] < 1$ ;  $p = 0.56$ ; right hemisphere:  $F[1,12] < 1$ ;  $p = 0.66$ ). Apart from this functionally defined ROI, we also analyzed the fMRI responses of the medial portion of Heschl's gyrus (HG) and the planum tem-

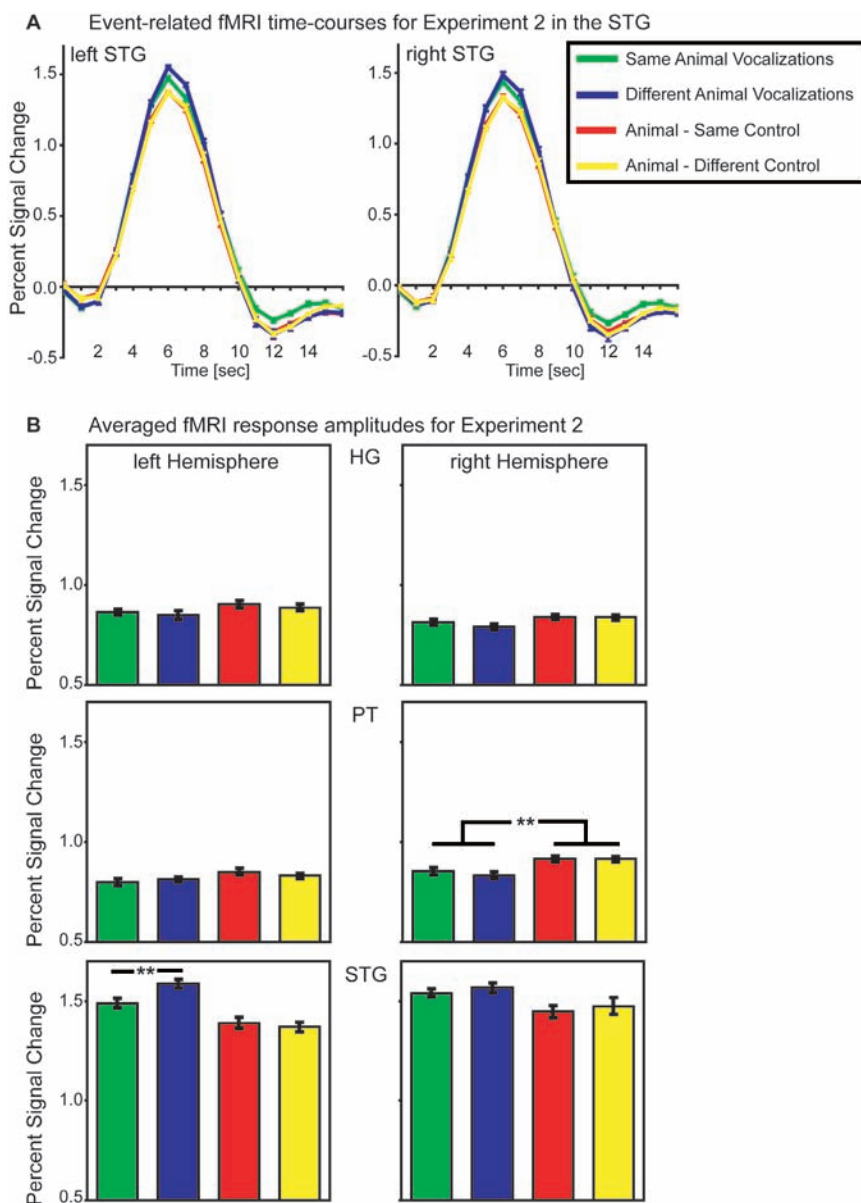


Figure 3. (A) Event-related time-courses of fMRI responses in the left and right superior temporal gyrus (STG) for Experiment 2 averaged across 13 subjects. (B) fMRI response amplitudes in Heschl's gyrus (HG), the planum temporale (PT) and the superior temporal gyrus (STG) for Experiment 2 averaged across 13 subjects. The error bars represent mean standard errors of the fMRI responses.

porale (PT). We observed significant differences between experimental conditions, in the right PT only ( $F[3,36] = 5.38$ ;  $p < 0.01$ ). More specifically, pairs with control stimuli showed a significantly stronger response compared to pairs of animal vocalizations only as determined by a two-way repeated measurements ANOVA ( $F[1,12] = 9.87$ ;  $p < 0.01$ ). Within the right PT, no adaptation effects were revealed for different versus same animal vocalizations ( $F[1,12] < 1$ ;  $p = 0.45$ ) or pairs of animal vocalizations and same versus different control ( $F[1,12] < 1$ ;  $p = 0.98$ ). As the fMRI responses for control stimuli were not larger compared to the animal vocalizations in Experiment 1 (paired t-test:  $t[12] = 0.94$ ;  $p = 0.37$ ), the differences in Experiment 2 are possibly based on the stimulus change, rather than the acoustic properties of single stimuli. We did not observe any significant experimental effects in HG or left PT (left HG:  $F[3,36] = 1.24$ ;  $p = 0.30$ ; right HG:  $F[3,36] = 1.53$ ;  $p = 0.22$ ; left PT:  $F[3,36] = 1.88$ ;  $p = 0.15$ ). There was no significant interaction between the factors Hemisphere (left/right PT) and Stimulus Type (animal vocalizations only/ animal vocalization and control) as tested with a two-way repeated measurements ANOVA ( $F[1,12] = 2.92$ ;  $p = 0.11$ ).

## Discussion

Employing two different paradigms, our study provides evidence for processing of animal vocalizations in the bilateral superior temporal gyrus (STG). Firstly, regions along bilateral STG showed a linear decrease of fMRI response strength with the degree of degradation of animal vocalizations. Secondly, within the left STG, fMRI adaptation effects were observed for same versus different animal vocalizations, suggesting a selective representation of these stimuli.

The bilateral pattern of fMRI responses to animal vocalizations is in line with studies that observed preferential fMRI responses to animal sounds versus tool sounds (Lewis et al., 2005) within the middle portions of the superior temporal gyrus (mSTG). Accordingly, areas within bilateral mSTG and the superior temporal sulcus (STS) have been implicated in the processing of human voices and vocalizations (Belin et al., 2000; Fecteau et al., 2004, 2005; Lewis et al., 2004). Furthermore, amongst other areas, the STG has been shown to respond more strongly to environmental sounds compared to white noise stimuli (Giraud and Price, 2001; Maeder et al., 2001). In a recent study, fMRI responses in the mSTG exhibited an increase for both recognized and unrecognized environmental sounds (Lewis et al., 2005). Thus, it has been proposed that the bilateral mSTG might represent an “intermediate” stage of auditory processing. That is, these areas are possibly involved in the encoding of acoustic structural attributes such as harmonic or phase-coupling content rather than higher level semantic information. In line with this assumption, areas in the superior temporal lobe have been associated with processing of temporal and spectral aspects of auditory stimuli. For example, recent fMRI studies have revealed overlapping areas within the lateral Heschl's gyrus and the planum temporale that are sensitive to both amplitude and frequency-modulated tones (Giraud et al., 2000; Hart et al., 2003). Intracranial recordings in human patients have demonstrated encoding of amplitude-modulation in the auditory cortex (Liegeois-Chauvel et al., 2004). Moreover, previous electrophysiological studies have suggested directional selectivity of neurons for frequency-modulated sweeps in

posterior non-primary auditory fields of the cat (Tian and Rauschecker, 1998). Possibly, the selective representation of the spectro-temporal dynamics of auditory stimuli constitutes the basis for the decoding of complex natural sounds such as animal vocalizations. Thus, similar adaptation effects within the STG might be expected for other stimulus classes with complex spectro-temporal dynamics such as human verbal and non-verbal vocalizations or other environmental sounds.

In this study, the bilateral STG exhibited stronger fMRI responses for animal vocalizations compared to temporally unstructured control stimuli. Adaptation effects, however, occurred within the left STG only. A possible explanation for the lateralization of adaptation effects might be the differential preference for timing parameters within left and right auditory cortex which has been reported in several neuroimaging studies. More specifically, the right superior temporal sulcus has been shown to respond preferentially to slowly modulated signals, while the left superior temporal sulcus showed preferences for faster spectro-temporal changes (Belin et al., 1998; Boemio et al., 2005). Similarly, within the human antero-lateral auditory belt, temporal auditory processing has been associated with the left-hemisphere and spectral processing with the right hemisphere (Zatorre and Belin, 2001; Schönwiesner et al., 2005). Thus, the observation of mainly left-hemispheric adaptation effects might be accounted for by the specific spectro-temporal dynamics of the employed animal vocalizations. That is, the spectro-temporal changes that distinguish between these stimuli possibly lie within a time range that is preferentially processed by the left STG. A previous study on fMRI repetition priming showed bilateral repetition-related reductions in the STS but right-lateralized reductions within the STG (Bergerbest et al., 2004). In contrast to the present experiments, this study employed general environmental sounds (animals and sounds produced by man-made objects) which might explain the obvious different hemispheric lateralization of repetition effects. Interestingly, a previous magnetoencephalographic (MEG) study has shown a bilateral distribution of evoked magnetic fields in response to mismatch of complex sounds (syllables, animal vocalizations and noise), but a left-hemispheric increase in gamma-band activity above 60Hz (Kaiser et al., 2002; Kaiser and Lutzenberger, 2005). Induced gamma-band activity has been associated with repetition-priming effects in EEG studies (Gruber and Müller, 2002; 2005) and has been suggested to play a complementary role compared to evoked potentials. Thus, the previously reported left-lateralized increase in gamma-band activity in response to pattern changes of complex sounds may be related to the left-lateralized adaptation effects observed in this study.

A different theory of auditory hemispheric lateralization holds that the left superior temporal lobe is preferentially involved in speech perception, while the right superior temporal lobe is more involved in environmental sound processing (Thierry et al., 2003). However, a recent review argued that processing of speech and environmental sounds often implicates overlapping areas, concluding that there are no speech-specific cortical regions, but that common correlates might mediate both processes (Price et al., 2005). Thus, the areas in the bilateral STG may be part of a multi-purpose processing stage for complex sounds of which the left STG selectively encodes the stimulus features that optimally distinguish between the different animal vocalizations used in the present study.

Interestingly, we observed significantly stronger fMRI responses in right PT for changes from animal vocalizations to degraded control stimuli compared to pairs of animal vocalizations only. In Experiment 1, we did not observe significant differences between stimuli, suggesting that the effects within right PT are due to the stimulus change rather than the properties of single stimuli. Possibly, in line with the hypothesis that lateralization of auditory processing is based on preferential tuning to temporal properties, right PT may be involved in the representation of spectro-temporal stimulus features with a long time constant. More specifically, a change from an animal vocalization to the control stimulus is a transition from a temporally structured stimulus to a sustained presentation of a broad-band stimulus. Such a transition might possibly be encoded in areas that preferentially respond to slower modulations of auditory stimuli. Nevertheless, future research is required to test systematically for effects of timing parameters of complex auditory stimuli on auditory fMRI adaptation within different auditory areas, employing a wide range of stimulus categories.

The use of an fMRI adaptation paradigm has been suggested to be able to provide additional information about the representation of information in the brain (Grill-Spector et al., 2006), complementing the results obtained by classical fMRI experiments that contrast responses to different stimulus categories. The latter approach relies on spatially separable representations of different stimulus categories. However, it cannot always be assumed that the spatial separation between neural populations encoding different categories is in the range of the resolution of fMRI, in particular in higher-order sensory areas. fMRI-adaptation experiments are able to overcome these restrictions and allow to systematically test selectivity or invariance to various forms of transformations such as the effects of visual viewpoint changes in the ventral visual stream (James et al., 2002) or effects of size and position changes in object-related visual areas (Grill-Spector et al., 1999). Within the visual domain, fMRI-adaptation effects have been replicated reliably in higher visual areas, whereas fMRI-adaptation effects in early visual cortex are currently under debate. More specifically, while some researchers have reported orientation-specific fMRI adaptation in V1 (Tootell et al., 1998) and motion direction-specific fMRI-adaptation in both lateral temporal-occipital and pericalcarine areas (Huettel and McCarthy, 2004), other groups failed to identify orientation-specific fMRI adaptation in V1 (Boynton et al., 2003; Murray et al., 2006). Thus, it has been suggested that not only adaptation effects, but also vascular mechanisms may contribute to the fMRI adaptation phenomenon. Recent studies (Sawamura et al., 2006) have employed an adaptation paradigm while acquiring single-cell recordings from macaque infero-temporal cortex for which fMRI adaptation has been observed previously (Sawamura et al., 2005). Although adaptation effects of single neurons were observed, (Sawamura et al., 2006) the tuning properties inferred from adaptation were sharper than the actual tuning of single neurons, suggesting that the link between adaptation and single cell tuning curves is not straightforward. For the auditory domain, neural adaptation effects have been obtained for pure tones in cat primary auditory cortex (Ulanovsky et al., 2003), and combined human fMRI and MEG studies have provided evidence for stimulus-specific adaptation effects both for pure tones and noise (Jääskeläinen et al., 2004) and vowels (Ahveninen et al., 2006) in non-primary auditory areas. Thus, employing adaptation paradigms might be a fruitful approach to investigate the representation of auditory information in the human brain. However, fur-

ther research is needed to establish the link between different methodologies, such as single cell recordings, human electrophysiology and brain imaging.

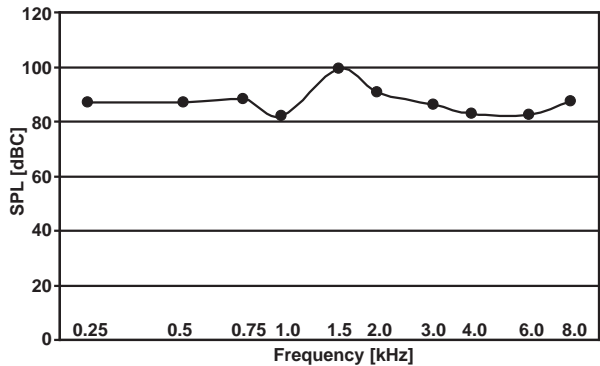
Conclusions

In summary, our study provides evidence for the involvement of the bilateral STG in the processing of animal vocalizations. Moreover, our data suggest selective representation of animal vocalizations within the left STG as determined by an event-related fMRI adaptation paradigm. Adaptation effects were predominantly observed for putatively non-primary auditory fields within the lateral STG rather than primary auditory cortex. The representation of animal vocalizations within this area is possibly mediated by the complex spectro-temporal dynamics of this stimulus class. These findings demonstrate that event-related fMRI adaptation is an appropriate method to test the selective representation of information in non-primary auditory fields. The investigation of cross-adaptation effects for different types of auditory features may lead to further insights into the organization principles of processing within auditory cortex.

Acknowledgements

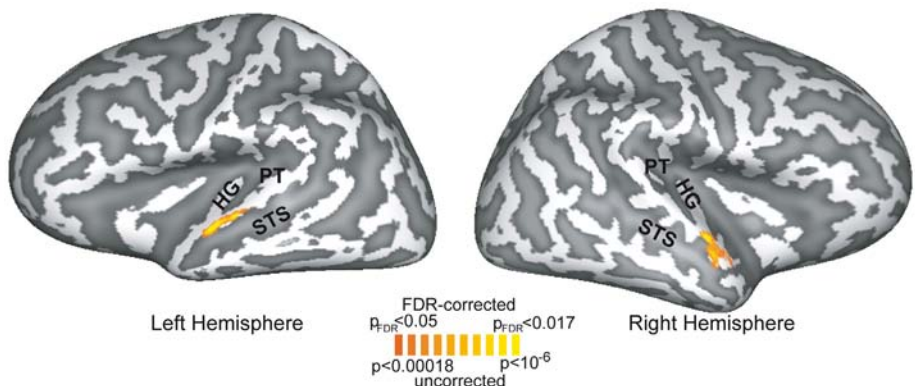
This study was supported by the Bundesministerium für Bildung und Forschung (Brain Imaging Center Frankfurt, DLR 01GO0203). The authors would like to thank Tim Wallenhorst for assistance with data acquisition.

Supplementary Figures

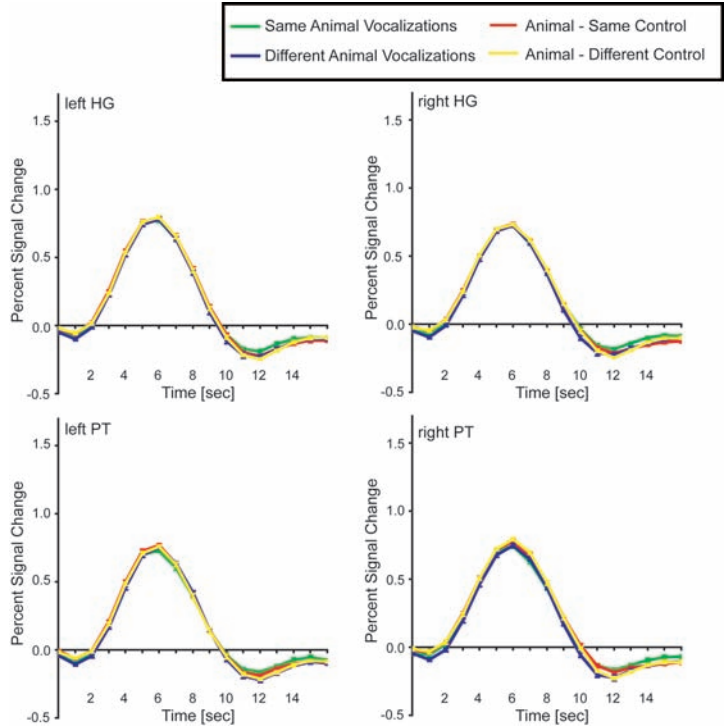


Suppl. Figure 1. Frequency response of the employed head phones (Commander XG, Resonance Technologies) as determined by measuring the sound pressure levels at 0.25 – 8 kHz. The frequency axis is logarithmic.





Suppl. Figure 2. Group data ( $n=13$ ) of the cortex-based general linear model analysis of Experiment 1 shown on an inflated 3D reconstruction of the cortical surface of a single subject which served as the target for cortex-based inter-subject alignment. Gyri are shown in light gray, sulci in dark gray. The activation map shows voxels that exhibited significantly ( $p_{FDR}<0.05$ , corrected for multiple comparisons using the False Discovery Rate approach described by Benjamini and Hochberg, 1995) stronger fMRI responses for intact versus degraded stimuli as determined by linear contrast analysis.



Suppl. Figure 3. Event-related time-courses of fMRI responses in the left and right HG and PT in Experiment 2 averaged across 13 subjects. The error bars represent mean standard errors of the fMRI responses.





# 3

## Probing category selectivity for environmental sounds in the human auditory brain

Earlier studies reported evidence suggesting distinct category-related auditory representations for environmental sounds such as animal vocalizations and tool sounds in superior and middle temporal regions of the temporal lobe. However, the degree of selectivity of these representations remains to be determined. The present study combined functional magnetic resonance imaging (fMRI) adaptation with a silent acquisition protocol to further investigate category-related auditory processing of environmental sounds. To this end, we consecutively presented pairs of sounds taken from the categories 'tool sounds' or 'animal vocalizations' with either the same or different identity/category. We examined the degree of selectivity as evidenced by adaptation effects to both or only one sound category in the course of whole-brain as well as functionally and anatomically constrained region of interest analyses. While most regions predominately in the temporal cortex showed an adaptation to both sound categories, particularly the left superior temporal gyrus (STG) and the left posterior middle temporal gyrus (pMTG) selectively adapted to animal vocalizations and tool sounds, respectively. However, the activation profiles of these regions differed with respect to the general responsiveness to sounds. While tool sounds still produced fMRI signals significantly different from fixation baseline in the STG, this was not the case for animal vocalizations in pMTG.

Consistent with the interpretation of STG as an intermediate auditory processing stage, this region might differentiate auditory stimuli into categories based on variations of physical stimulus properties. However, processing in left pMTG seems to be even more restricted to action-related sounds of man-made objects.

Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., and Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia*, 46, 2776-2786.

## Introduction

Sound identity processing has been found to involve temporal and frontal cortical regions in non-human primates (Romanski et al., 1999; Rauschecker and Tian, 2000; Romanski and Goldman-Rakic, 2002) and humans (Clarke et al., 2000; Alain et al., 2001; Kaiser et al., 2003; Arnott et al., 2004). The proposed functional specialization of temporo-frontal areas for auditory recognition has been refined by functional magnetic resonance imaging (fMRI) studies using auditory stimuli with different degrees of complexity like frequency- and amplitude-modulated sounds (Hall et al., 2002), speech (Belin et al., 2000; Binder et al., 2000; Christoffels et al., 2007), music (Patterson et al., 2002), or broad sets of object-related environmental sounds (Lewis et al., 2004; Zatorre et al., 2004). Environmental sounds can be further differentiated into categories, e.g. a distinction can be made between sounds from an animate versus inanimate source. Previous neuroimaging studies have found category-specific activation patterns in temporal cortex distinguishing between animate (animal vocalizations) and inanimate (tool sounds) categories (Lewis et al., 2005; Engelien et al., 2006; Murray et al., 2006). More specifically, Lewis et al. (2005) have reported activations in middle portions of bilateral superior temporal cortex (STG) for animal vocalizations compared to tool sounds, whereas the reverse contrast revealed a widespread activation of left-hemispheric regions including the posterior middle temporal gyrus (pMTG) and additional parietal and frontal areas. These reports of differential category-related responses are to some extent in line with neuropsychological findings of distinct lesion foci for conceptual knowledge pertaining to animals and tools (see e.g. Tranel et al., 1997). More specifically, regions in the proximity of the left pMTG have repeatedly been associated in lesion and neuroimaging studies (e.g. Whatmough et al., 2002; Tranel et al., 2003b) with tool and action processing both on a perceptual and on a recognition-related level. Further evidence suggests that additional regions of the left inferotemporal (IT) cortex might then be engaged in the context of naming tasks (Damasio et al., 2001; Tranel et al., 2003a). Animal and tool stimuli presented both in the auditory and visual sensory modality seem to activate these left IT regions to a similar degree, but again with differential sites with anterior-ventral and posterior-lateral foci for animal and tool stimuli, respectively (Tranel et al., 2005).

The present study aimed at further elucidating cortical preferences during earlier perceptual and recognition-related processing stages for animate versus inanimate sounds by applying an fMRI-adaptation approach (Grill-Spector and Malach, 2001; Grill-Spector et al., 2006). It has been used successfully in the visual domain to elucidate the neural correlates of visual shape processing (Kourtzi et al., 2003; Altmann et al., 2004) and object recognition (e.g. Vuilleumier et al., 2002). In contrast to the visual domain, relatively few studies have employed this method to assess differential sensitivities of cortical regions to auditory stimulus properties (Belin and Zatorre, 2003; Bergerbest et al., 2004; Dehaene-Lambertz et al., 2006). For instance, we found adaptation effects to repeated animal vocalizations in left STG that were absent for the successive presentation of an animal vocalization and a temporally degraded version of the same sound, demonstrating that the effect was not attributable to a mere adaptation to stimuli with similar spectral features (Altmann et al., 2007).

An important advantage of fMRI adaptation is its capability to extend findings which are based on the simple subtraction of fMRI responses to different stimuli. In the present context of auditory category selectivity, fMRI adaptation allowed us to determine in which cortical regions putative category-related preferences translate into actual corresponding tool- or animal-selective adaptation effects. In contrast, if these regions showed an adaptation to the repetition of auditory stimuli irrespective of their category, possibly because of adaptation to features like complex spectro-temporal dynamics (see e.g. Nelken, 2004) common to both categories, they might be regarded as nonselective. To address this question we compared fMRI responses to presentations of pairs of 1) identical animal vocalizations, 2) different animal vocalizations, 3) identical sounds of man-made tools, 4) different tool sounds, and 5) an animal vocalization and a tool sound.

The central research question of the present study can be stated as the question of which cortical regions do actually exhibit category-related selectivity as determined by fMRI adaptation. More explicitly, we expected adaptation effects selective to animal vocalizations in the left STG (Altmann et al., 2007). However, given its role as an “intermediate” auditory processing stage we also expected to find a general responsiveness (fMRI signal different from baseline) to sounds from the other category, but no corresponding adaptation effects for the tool sounds. On the other hand, previous studies have suggested a preference for tool sounds (compared to animal vocalizations) in several regions of the temporal, parietal and frontal cortices (Lewis et al., 2005). Within this network, we intended to identify regions which show an additional selective adaptation effect for tool sounds. Given its prominent role for semantic processing (see e.g. Noppeney et al., 2005) even beyond the auditory domain (see e.g. Beauchamp et al., 2004), we expected to find tool-selective adaptation effects particularly in PMTG.

The finding of category-selective fMRI-adaptation effects in these or possibly other cortical regions would therefore provide complementary evidence for their importance in the processing of these stimulus categories.

## Materials and Methods

### Subjects

Fifteen healthy, right-handed volunteers participated in the study. Handedness was evaluated by self-report. All subjects had normal hearing abilities and gave their informed consent to participate in the study. The experiments were performed in accordance with the ethical standards laid down in the 1964 declaration of Helsinki. One subject had to be excluded from the analysis because of excessive head motion. The mean age of the remaining 14 participants was 27.1 years (range 21-34; 9 males).

### Stimuli

To investigate auditory category selectivity, we employed two different categories of environmental sounds, namely animal vocalizations and tool sounds (Fig. 1A). Fifteen stimuli per category (Table 1) were taken from a collection of professionally recorded sounds (Sound Ideas, Richmond Hill, Ontario, Canada). The sound stimuli were digitized with a sampling rate of 22050 Hz. The sound duration did not differ between cate-

Table 1. List of stimuli

Stimulus (type)	Recognized (number of subjects)	Confidence (Score 1-7; 7 maximum)	Prototypicality (Score 1-7; 7 maximum)
<i>Animal vocalizations</i>			
Cat	6/6	7.0	6.7
Chicken	6/6	6.8	6.7
Chimp	4/6	6.5	5.7
Cow	6/6	7.0	6.2
Crow	6/6	6.2	5.5
Dog	6/6	7.0	6.8
Donkey	6/6	6.8	6.8
Duck	6/6	6.0	5.5
Elephant	6/6	7.0	6.5
Frog	6/6	7.0	6.8
Horse	6/6	7.0	6.8
Owl	6/6	6.5	6.2
Seal	3/6	4.3	4.3
Sheep	6/6	7.0	6.8
Tiger	4/6	4.7	4.5
Mean +/- SEM	5.5 +/-0.75	6.5 +/-0.62	6.1 +/-0.68
<i>Tool sounds</i>			
Bikebell	6/6	7.0	7.0
Camera	5/6	6.8	6.0
Chainsaw	5/6	6.7	6.8
Chisel (hammered)	6/6	6.0	5.7
Drill	6/6	5.3	4.7
File	6/6	5.5	5.8
Hammer (hitting nail)	6/6	5.7	5.3
Hammer (hitting wood)	6/6	6.5	6.5
Jackhammer	6/6	7.0	6.8
Ratchet	5/6	6.2	5.8
Sanding	5/6	4.3	4.2
Saw	6/6	6.3	6.2
Scissor (cutting paper)	3/6	3.5	2.2
Stapler	5/6	5.7	5.5
Typewriter	5/6	5.2	4.0
Mean +/- SEM	5.27 +/- 0.68	5.8 +/- 0.77	5.5 +/- 0.96

&lt;&lt;&lt;

Recognition scores with confidence ratings obtained from six subjects, who did not participate in the fMRI sessions, for the auditory stimuli of the two categories. Subjects were also instructed to rate the prototypicality of the stimuli, i.e. how well the chosen exemplar represented the “typical” sound of the respective object.

gories ( $t[14] = -.562$ ;  $p > 0.1$ , paired t-test). It was on average 1048 ms for animal stimuli (standard deviation 156 ms, range: 734–1386 ms) and 1085 for tool stimuli (standard deviation 143 ms, range: 771–1342 ms).

Six additional subjects who did not participate in the fMRI experiments were instructed to recognize the stimuli of the two categories and to provide a confidence rating for their decisions on a Likert scale ranging from 1 (“not confident”) to 7 (“perfectly confident”). Furthermore, each stimulus was rated with respect to its “prototypicality”, i.e. how well the given sound exemplar represented the object, again on a Likert scale ranging from 1 (“not prototypical”) to 7 (“perfectly prototypical”). The scores for both stimulus categories were very similar (Table 1) and on a comparable level with ratings reported in earlier studies (Murray et al., 2006; Altmann et al., 2007). Thus, differential experimental effects for the two stimulus categories were unlikely to be attributable to these parameters.

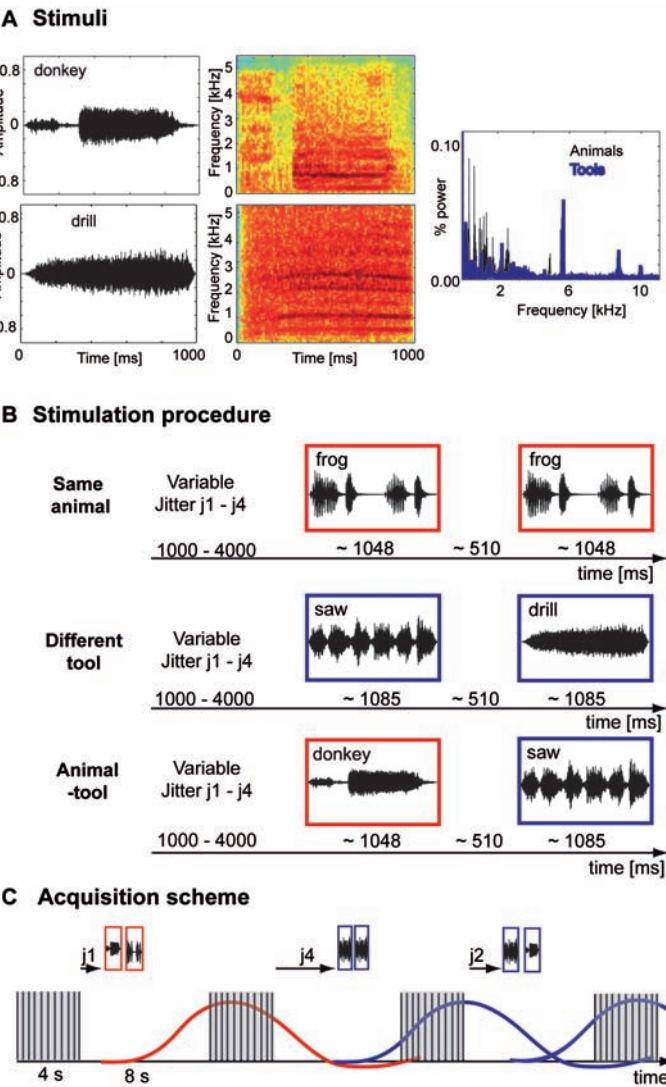
All sounds were equalized with regard to their root mean square sound intensity. Stimuli were presented with MR-compatible head phones (Resonance Technology, Northridge, USA) using wave guide tubes for sound transmission. The headphones of the MRI audio system contained fluid- and vinyl-sealed cushions to attenuate ambient noise up to 30 dB. For each subject the applied output volume for the stimulus presentation was adjusted individually to a comfortable level resulting in an average sound pressure level of 82 dB(C) (animals: 82.9 dB(C), tools: 81.5 dB(C)). The frequency response of the head phones ensured reliable presentation of the employed stimuli as described in a previous study (Altmann et al., 2007).

The averaged power spectrum for stimuli of the categories used in the present study (Fig. 1A) revealed that animal vocalizations typically contained a greater relative power in the range between ~ 1000 to 6000 Hz. In contrast, tool sounds had a relatively uniform power distribution over the complete spectrum with local maxima at ~ 2000, 6000 and 10000 Hz. The mean harmonic-to-noise ratio of our animal stimuli as determined by the software package Praat (<http://www.fon.hum.uva.nl/praat/>) was 6.18 dB as compared to 0.36 dB for the tool sounds. These values are in a similar range to previously reported work on auditory object recognition (Lewis et al., 2005).

## Procedure

Subjects were administered four runs of an event-related fMRI-adaptation experiment and one run of a high-resolution T1-weighted MP-RAGE sequence as an anatomical reference.

Essentially, the rationale for an fMRI-adaptation paradigm is its capability to identify subpopulations of neurons sensitive to certain stimulus features even when these are embedded in a larger cortical network. For instance, in a conventional fMRI experiment a cortical region *R* might react quite uniformly to the presentation of audi-



tory stimuli differing from each other along a certain variable of interest, for example, amplitude-modulation (AM) rate. Based on negative results from simple subtraction of responses to stimuli with different AM rates, *R* might be regarded as insensitive to variations in this property. In contrast, an fMRI-adaptation paradigm utilizes the well-established observation of a decreased fMRI signal (Grill-Spector and Malach, 2001) for blocks (Belin and Zatorre, 2003; Bergerbest et al., 2004) or successively-presented pairs of repeated sounds (Altmann et al, 2007). For the latter type of experimental protocol, these fMRI-responses to “same” conditions are compared to responses to pairs of stimuli varying the respective property of interest (“different” conditions). Thus, a stronger fMRI signal in region *R* evoked by stimulus pairs with different AM rates when compared to the repetition of the same stimuli indicates sensitivity to AM rates. Note, however, that these signal differences are solely induced by the immediate effect the first stimulus exerts on the second one.

Following this general reasoning the event-related fMRI experiment contained five different experimental conditions (Fig. 1B): 1) two identical animal vocalizations (“animal-same”), 2) two different animal vocalizations (“animal-different”), 3) two identical tool sounds (“tool-same”) 4) two different tool sounds (“tool-different”) and 5) an animal vocalization or tool sound paired with a stimulus of the other category (“animal and tool”). Each trial lasted for 12 seconds and started with a variable non-stimulus delay of 1, 2, 3 or 4 seconds. This delay served as a jitter to maximize the sampling of different subsections of the blood oxygen level-dependent (BOLD) signal and to ensure efficient parameter estimation of the hemodynamic response function (Dale et al., 1999). Next, either an animal or tool sound was presented with a mean duration of 1048 ms or 1085 ms, respectively. After a short interstimulus interval of 510 ms on average, this adapting stimulus *S1* was either followed by the same sound (producing the two “same” conditions) or a different sound *S2* either taken from the same (producing the within-category “different” conditions) or the other category (producing the across-category “different”-condition). Due to our randomization procedure each stimulus was equally likely to be presented as *S1* and *S2*.

Each experimental run of the fMRI experiment lasted 10 min and 12 s and was comprised of 42 pairs of auditory stimuli. Within a single run, 7 trials per condition were presented, intermixed with 7 trials without auditory stimulation (baseline). For each subject, four experimental runs were conducted, resulting in 28 trials per experimental conditions. Similar to previous studies (Altmann et al., 2004), the order of presentation was counterbalanced so that trials from each condition were preceded (1 trial back) equally often by trials from each of the other conditions.

Throughout the experiment, subjects were instructed to keep their eyes open and fixate a white cross on a black background back-projected by an LCD-projector through a mirror fixed to the head coil. Subjects were instructed to press a button whenever a specific target sound was presented. This target sound, for instance the cat vocalization, was individually assigned to each participant prior to the first experimental run. The category of the target (animal vocalization or tool sound) was balanced across subjects. To avoid the contamination with task-related motor activity, the respective target trials were discarded from further analysis of the imaging data. Target stimuli were presented 7 times per experimental run resulting in no substantial loss of experimental trials. This task provided an efficient attentional control compared to



passive listening (Engelien et al., 2006). Because of technical problems behavioral data were not recorded for one subject. For the remainder of the sample the rate of correct responses was 98.6 percent. For the whole experiment, subjects were instructed to use only one hand for responding. The assignment of the left or the right hand was counterbalanced across subjects.

### Imaging and Data Analysis

Imaging was conducted using a 3T ALLEGRA (Siemens, Erlangen, Germany) scanner at the Brain Imaging Center, Frankfurt am Main. Several recent studies (e.g. Gaab et al., 2007a,b; Schmidt et al., 2008) have demonstrated an influence of ambient scanner noise during continuous scanning procedures on cortical responses particularly in regions of the superior temporal plane. Such an acoustical stimulation due to gradient switching might cause a contamination of hemodynamic responses induced by stimulus presentation (see e.g. Hall et al., 1999) and possibly reduces the ability of subjects to correctly identify stimuli. We therefore employed an interleaved silent steady state echo planar imaging (EPI) sequence, similar to recent auditory fMRI studies (Schwarzbauer et al., 2006), to reduce these influences of scanner noise. During the silent period of the sequence, the EPI readout was omitted while slice selection was continued with same repetition time (TR), excitation pulses and gradients as compared to the acquisition period in order to keep the longitudinal magnetization in a steady state. For slice selection, silent gradients with long ramp times (whisper mode) were employed generally, with only minor influence on imaging time. Each acquisition block lasted 12 seconds and consisted of a silent period of 8 seconds during which the acoustic stimulus was presented and a data acquisition period of 4 seconds. The following EPI parameters were used: TR = 2 s, TE = 30 ms. Thirty-three axial slices (3 mm thickness with  $3.00 \times 3.00 \text{ mm}^2$  in-plane resolution, inter-slice gap: 0.6 mm) were collected with a bird-cage head coil. The field of view was  $19.2 \text{ cm} \times 19.2 \text{ cm}$  with an in-plane resolution of  $64 \times 64$  pixels. As an anatomical reference we acquired 3D volume scans by using a magnetization-prepared rapid-acquisition gradient echo sequence (MP-RAGE; 160 slices; TR: 2300 ms).

The presentation of stimulus pairs was jittered in one-second steps between one and four seconds within this silent period (Fig. 1C). Accordingly, the hemodynamic response was shifted differently with respect to the image-acquisition period resulting in a sampling of several subsections of the BOLD signal. This was used for more efficient parameter estimation of the hemodynamic response function.

Functional data were pre-processed and analyzed using the Brainvoyager QX (Brain Innovation, Maastricht, Netherlands) software package. Pre-processing included head movement correction and linear de-trending. The 2D functional images were co-registered with the 3D anatomical data, and both 2D and 3D data were spatially normalized into the Talairach and Tournoux stereotactic coordinate system (Talairach and Tournoux, 1988). Functional 2D data were spatially filtered employing a Gaussian kernel of 8 mm (full width at half maximum; FWHM).

We computed a general linear model (GLM) to estimate the beta weights that model the hemodynamic responses as a combination of two gamma functions (Friston et al., 1995; Friston et al., 1998). A whole-brain random-effects (RFX) GLM was calculated which included the following nine predictors: 1) five predictors for the experimental

conditions involving auditory stimulation, 2) one predictor for the fixation condition, 3) one predictor for the target condition involving task-related motor activation 4) one predictor for the auditory stimulation resulting from the scanner noise and 5) one predictor for residual T1-related signal decay. The GLM was masked with respect to those voxels which generally showed a stronger activation to all auditory conditions compared to fixation using a threshold of  $p < 0.05$  (uncorrected). This liberal threshold was chosen to include as many auditorily responsive voxels as possible and to avoid missing significant effects in further analysis steps in regions which could not be covered by more restricted masks. Beta weights obtained from this RFX GLM were entered into a two-way whole-brain analysis of variance (ANOVA) with the factors “category” (animals/tools) and “repetition” (same/different).

In a first analysis step we used this ANOVA to search for significant main effects of “category” pooling across the “same” and “different” conditions to identify those voxels which were preferentially activated for one of the categories. The results of this contrast (Fig. 2) were projected onto axial slices of one subject of our sample. In a region of interest (ROI) analysis we evaluated if any of these functional clusters showed additional adaptation effects separately (selectively) for the two auditory categories by contrasting each “different” condition to the respective “same” condition. For instance, “animal-different” > “animal-same” identified those cortical regions which were particularly sensitive to differences between animal vocalizations.

In a second analysis step we then evaluated whole-brain adaptation effects to identify additional regions which were not revealed by the previous contrast. To this end, we contrasted each “same”- and “different” condition separately for each category on a whole-brain level. Functional clusters which were revealed by these contrasts served as ROIs to further assess their responses to other experimental conditions. More precisely, we were interested if regions that showed significant adaptation to the repetition of stimuli of one category also significantly adapted to repeated stimuli of the other category (“non-selective adaptation”). Additionally, we searched within these ROIs for main effects of the respective stimulus category and for significant differences between the “animal and tool” condition compared to the mean of the two “different” conditions. The rationale for the latter approach was that the “animal and tool” condition involved not only spectro-temporally different stimuli but also a change of stimulus category. Thus, it could serve as a control condition which enabled the evaluation of responses both to spectro-temporal and category-related changes. The response profiles of the ROIs in Figure 3 are reported after the subtraction of the fixation condition. Additionally, for these profiles as well as for the results of the contrasts, the standard-error bars of the beta weights were corrected for inter-subject effects similar to previous studies (Altmann et al., 2004). In particular, the fMRI responses were calculated individually for each subject by subtracting the mean percent signal change for all conditions within this subject from the mean percent signal change for each condition and adding the mean percent signal change for all the conditions across subjects.

Finally, to further relate our results to the underlying anatomy of the human auditory brain, we conducted a second analysis using anatomical constraints derived from single-subject data. The aim of this analysis was to minimize potential confounds due to a profound inter-subject variability of the exact anatomical locations of certain auditory regions which might have caused a lack of overlap of activation on a group level.

To this end, we first created a three-dimensional reconstruction of both cortical hemispheres of each subject. Based on these reconstructions we then individually identified five anatomical structures commonly reported in studies on auditory perception in humans (Fig. 4). Heschl's gyrus (HG) was identified based on its posterior-medial to anterior-lateral progression within the lateral fissure. If two or three gyri were present, all of them together with the sulcus intermedialis were included in this ROI. The planum temporale (PT) was defined as a region from posterior to HG to the end of the Sylvian fissure marked by the ascending and descending ramus (see Hall, 2005). Anterior to HG we identified the planum polare (PP) covering the anterior portion of the superior temporal plane. PT and PP were both medially bordered by the insular cortex. Finally, ROIs were defined on the superior temporal gyrus (STG) with a common border on the level of HG. The projected anterior and posterior borders of PP and PT, respectively, also served as the borders for the ROI covering the anterior (aSTG) and posterior STG (pSTG). Besides these anatomical criteria we also restricted the voxels included in the analysis on the basis of functional criteria. For each subject we computed a GLM with the same predictors as mentioned before for the group data. We then contrasted all auditory conditions with fixation and applied a threshold of  $p < 0.001$ . These activation maps were used as masks to include only those voxels in the ROI analysis which showed significant responses to auditory stimulation. This aimed at avoiding averaging across neural populations with a frequency tuning beyond the range of the employed stimuli (Altmann et al., 2007). Finally, ROI analysis was further restricted in HG and PT to those voxels that fell within the 90% boundaries of the probability map for HG (Rademacher et al., 2001) and the 95% boundaries for PT (Westbury et al., 1999).

Based on these individually identified voxels we then compared the percent signal change values of all subjects across conditions by using a two-way ANOVA for repeated measurements with the factors "sound category" (animal/tool) and "repetition" (same/different). The degrees of freedom were corrected with the Greenhouse-Geisser method whenever appropriate. Furthermore, we also computed t-tests between the "same" and "different" conditions of each category to evaluate if the respective regions show an adaptation only for one or both sound categories.

## Results

### Whole-brain main effects for animal vocalizations and tool sounds

To localize cortical areas that were involved preferentially in the processing of animal vocalizations and tool sounds, respectively, we computed a random-effects general linear model (GLM) to characterize fMRI responses common to all subjects ( $n=14$ ). In a first analysis step those regions were identified that showed a main effect of stimulus category, pooling over the respective "same" and "different" conditions. This linear contrast analysis (Fig. 2) revealed bilateral areas of the superior temporal gyrus (STG), the left inferior frontal gyrus, the right superior parietal lobule, and the right anterior insula that exhibited significantly ( $q(\text{FDR}) < 0.05$ ) stronger fMRI responses to animal vocalizations as compared to tool sounds. For the reverse contrast, i.e. tool sounds compared to animal vocalizations, we observed significantly stronger fMRI responses involving the bilateral posterior middle temporal gyrus (pMTG) and adjacent regions of the su-

perior temporal sulcus (STS) and STG in the right hemisphere, the left postcentral gyrus /insula, middle frontal gyrus and bilateral regions on anterior portions of the superior temporal plane, corresponding to the planum polare (PP). Table 2 summarizes the activation extent and exact location (Talairach coordinates) of the observed cortical regions.

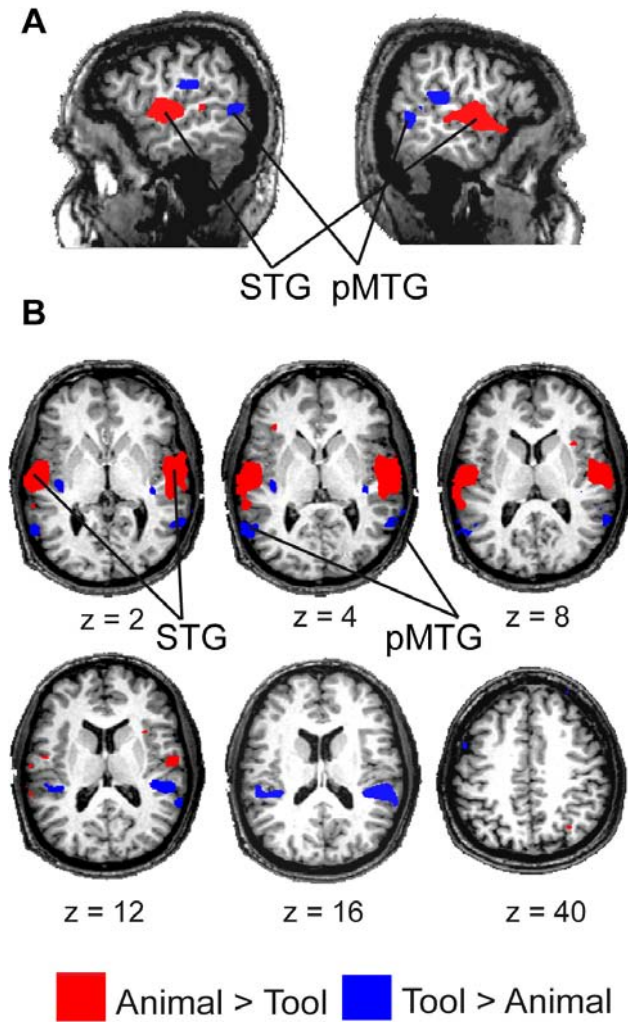


Figure 2. Main effects based on a whole-brain group analysis. (A) Group data ( $n=14$ ) derived from the general linear model analysis of the experiment are shown on two sagittal and (B) six different axial slices of the anatomical dataset of one representative subject. The activation maps display voxels that exhibited significantly ( $q(\text{FDR}) < 0.05$ ) stronger fMRI responses for animal vocalizations (red) than tool sounds and for the opposite contrast (blue).

Table 2. Coordinates and number of significant voxels in regions showing a main effect of category

Region of interest	x	y	z	Number of voxels
<i>Animals &gt; Tools</i>				
Left STG	-57	-15	5	5925
Left inferior frontal gyrus	-36	28	5	51
Right STG	55	-8	4	7004
Right anterior insula	33	15	10	129
Right superior parietal	32	-57	42	77
<i>Tools &gt; Animals</i>				
Left PP	-38	-18	1	556
Left STG /Insula	-33	0	-10	42
Left pMTG	-57	-55	5	924
Left postcentral gyrus / Insula	-48	-29	18	1384
Left middle frontal gyrus	-53	10	40	52
Right anterior PP	37	-7	-9	531
Right PP	37	-23	3	147
Right STG / pMTG	52	-35	13	3675

Regions with a main effect of category as revealed by an ANOVA based on a RFX GLM. Talairach coordinates in millimeter for the group-based whole-brain analysis of adaptation effects. The number of voxels is based on the resolution of the anatomical data set, i.e. 1x1x1 mm. Abbreviations: HG = Heschl's gyrus, pMTG= Posterior middle temporal gyrus, PP = Planum polare, PT = Planum temporale STG = Superior temporal gyrus.

### ROI analysis based on regions showing main effects for sound category

For a first analysis of adaptation effects, regions with a significant main effect were used as functionally defined ROIs. Random-effects GLMs were computed in these ROIs to determine possible adaptation effects by comparing the "different" and "same" conditions separately for each category. Two regions which were determined by a main effect for animal vocalizations showed significant adaptation, that is significant effects between the respective "same"- and "different"- conditions: the left STG for animal vocalizations ( $t(13)=2.690$ ;  $p<0.05$ ) and the right anterior insula for tool sounds ( $t(13)=3.332$ ,  $p<0.05$ ). In contrast, two regions as determined by a main effect for tool sounds showed significant adaptation to tool sounds only: the left pMTG ( $t(13)=2.437$ ;  $p<0.05$ ) and the left PP ( $t(13)=2.791$ ;  $p<0.05$ ).

### Whole-brain adaptation effects for animal vocalizations and tool sounds

To identify additional regions with adaptation effects, we calculated contrasts on a whole-brain level (separately for each category) between the respective "same"- and

"different"- conditions. Significant effects ( $t(13)=3.38$ ;  $p<0.005$ , uncorrected) for the contrast "animal-different" > "animal-same" (Fig.3 A) were found bilaterally in Heschl's gyrus (HG; extending into the STG in the right hemisphere), in a region located anterior laterally to HG on the STG, in posterior STG, and in regions located in the left anterior insula and the right PP. The contrast "tool-different" > "tool-same" identified - apart from a small cluster on the right HG - a strongly left-lateralized network comprising the left anterior insula, two subregions of the PP, the left HG / PT and two separate clusters in pMTG (see Table 3 for information concerning the exact spatial extent and the Talairach coordinates of these functional clusters.)

### **ROI analysis in regions with significant adaptation effects for one category**

For an analysis of the response profiles of these previously identified regions we again conducted random effect ROI-GLMs and extracted the estimated beta weights for the different conditions. Moreover, t-tests were calculated to determine adaptation for the respective other sound category and to search for additional effects in these regions. These computations (Fig. 3B) revealed that most of the regions identified before as adapting to one stimulus category, also showed adaptation to the other category. This was in accordance with our definition of non-selective adaptation. In the ROIs based on animal adaptation effects, the left anterior STG was the sole exception. In this region the repetition of tool sounds ("tool-same") did not produce a significantly reduced response when compared to the presentation of different tool sounds ( $t(13)=1.108$ ,  $p=0.28$ ). However, this region was still responsive to tool sounds compared to silent intervals since the beta weights for these conditions were still significantly different from baseline. In line with the results from the first analysis step, a significant main effect for animal vocalizations compared to tool sounds was also found only in this region ( $t(13)=2.432$ ,  $p<0.05$ ). Finally, voxels in this ROI were significantly responsive to the animal and tool"-condition when compared to the mean of the two "same" conditions ( $t(13)=2.174$ ,  $p<0.05$ ), but not to the mean of the "different" conditions ( $t(13)=-0.390$ ,  $p=0.7$ ).

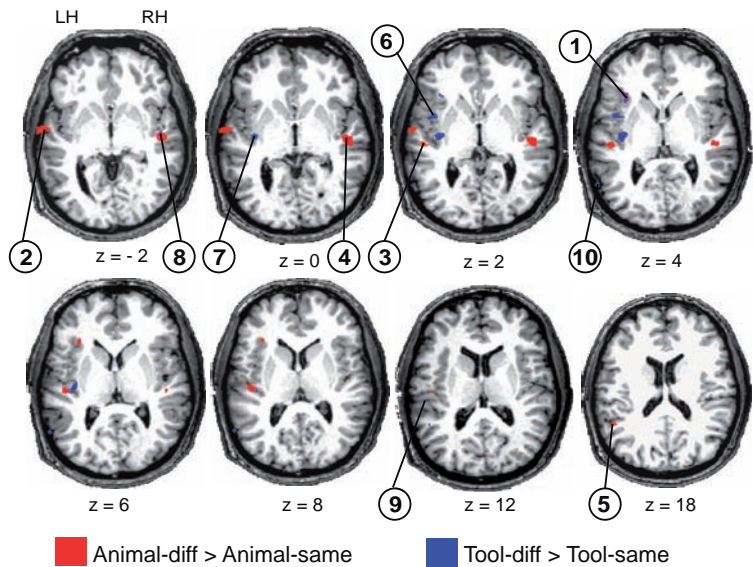
ROIs determined by the whole-brain adaptation effect for tool sounds in most cases also adapted to the animal vocalizations, with three exceptions: the left PP (animal adaptation effect:  $t(13)=1.738$ ,  $p=0.11$ ), the left HG/ STG ( $t(13)=0.050$ ,  $p=0.96$ ) and the left pMTG ( $t(13)=0.347$ ,  $p=0.73$ ). Furthermore, the left pMTG is the only ROI with additional significant effects, that is, a main effect of tool sounds compared to animal vocalizations ( $t(13)=2.75$ ,  $p<0.05$ ). Note, however, that the responses to the presentation of animal vocalizations were not significantly different from silent baseline.

### **Anatomically constrained analysis of main- and adaptation effects**

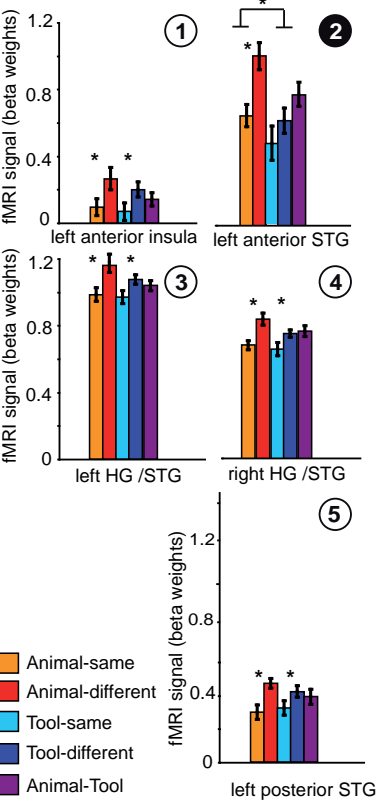
A final analysis step focused on the anatomical organization of the superior temporal plane in human cortex. By defining 10 regions of interest (HG, PP, PT, anterior and posterior STG in both hemispheres; see Fig. 4A) for each subject in our sample, we intended to relate adaptation effects more precisely to anatomical substrates (see Petkov et al., 2004 for a similar approach) and to reduce possible confounds for the group-based analysis caused by a substantial degree of anatomical intersubject variability even on very early cortical levels such as HG. This anatomically constrained analysis revealed for each ROI a significant main effect of repetition suggesting a general responsiveness



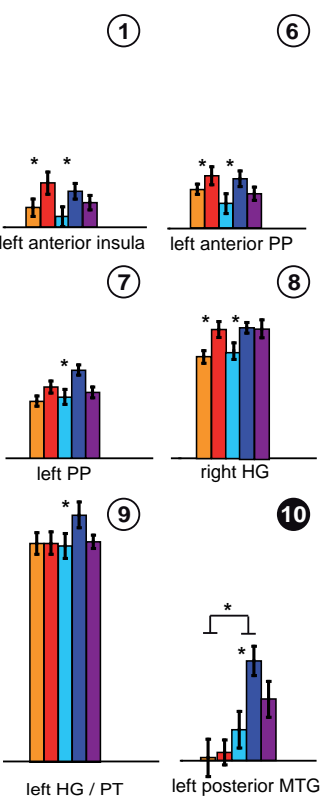
**A Whole-brain adaptation effects**



**B Animal adaptation effects**



**Tool adaptation effects**



<<<

Figure 3. Adaptation effects based on a whole-brain group analysis. (A) Eight axial slices of the same representative subject as in Figure 2 showing voxels with significant ( $p < 0.005$ , uncorrected) adaptation effects as determined by the group-based contrast “animal-different” > “animal-same” (shown in red) and “tool-different > tool-same” (shown in blue). (B) Response profiles (beta weights) of those cortical regions shown in A. Regions with significant adaptation effects for animal vocalizations (left panel) and tool sounds (right panel) were evaluated with regard to adaptation effects ( $p < 0.05$  as denoted by asterisks) for the respective other category as well as for main effects of stimulus category. Colors code for the respective experimental conditions: orange, repetition of the same animal vocalization (animal-same); red, presentation of two different animal vocalizations (animal-different); light blue, repetition of the same tool sound (tool-same); dark blue, presentation of two different tool sounds (tool-different); violet, presentation of an animal vocalization and a tool sound (animal and tool). The responses are reported after subtraction of the beta weights corresponding to the fixation condition on a single-subject basis and error bars represent mean standard errors of the fMRI responses corrected for inter-subject effects. See Methods section for details. Cortical regions with the most relevant experimental effects were denoted with a white number on a black ground.

Table 3. Coordinates of regions and number of significant voxels showing a whole-brain adaptation effect

Region of interest	x	y	z	Number of voxels
<i>Animal-different &gt; Animal-same</i>				
Left HG	-44	-19	8	258
Left anterior STG	-61	-7	-1	395
Left posterior STG	-49	-45	18	84
Left anterior Insula	-33	23	6	121
Right HG /STG	47	-17	1	544
Right PP	39	-14	6	20
<i>Tool-different &gt; Tool-same</i>				
Left HG /PT	-52	-25	13	26
Left PP	-36	-15	4	365
Left anterior PP	-41	4	3	163
Left anterior Insula	-34	21	4	104
Left pMTG	-59	-58	6	80
Left STS/pMTG	-47	-47	12	20
Right HG	45	-12	-2	29

Regions with a main effect of category as revealed by an ANOVA based on a RFX GLM. Contrasts of “different” and “same” conditions were calculated separately for each category. The same conventions apply as in Table 2.



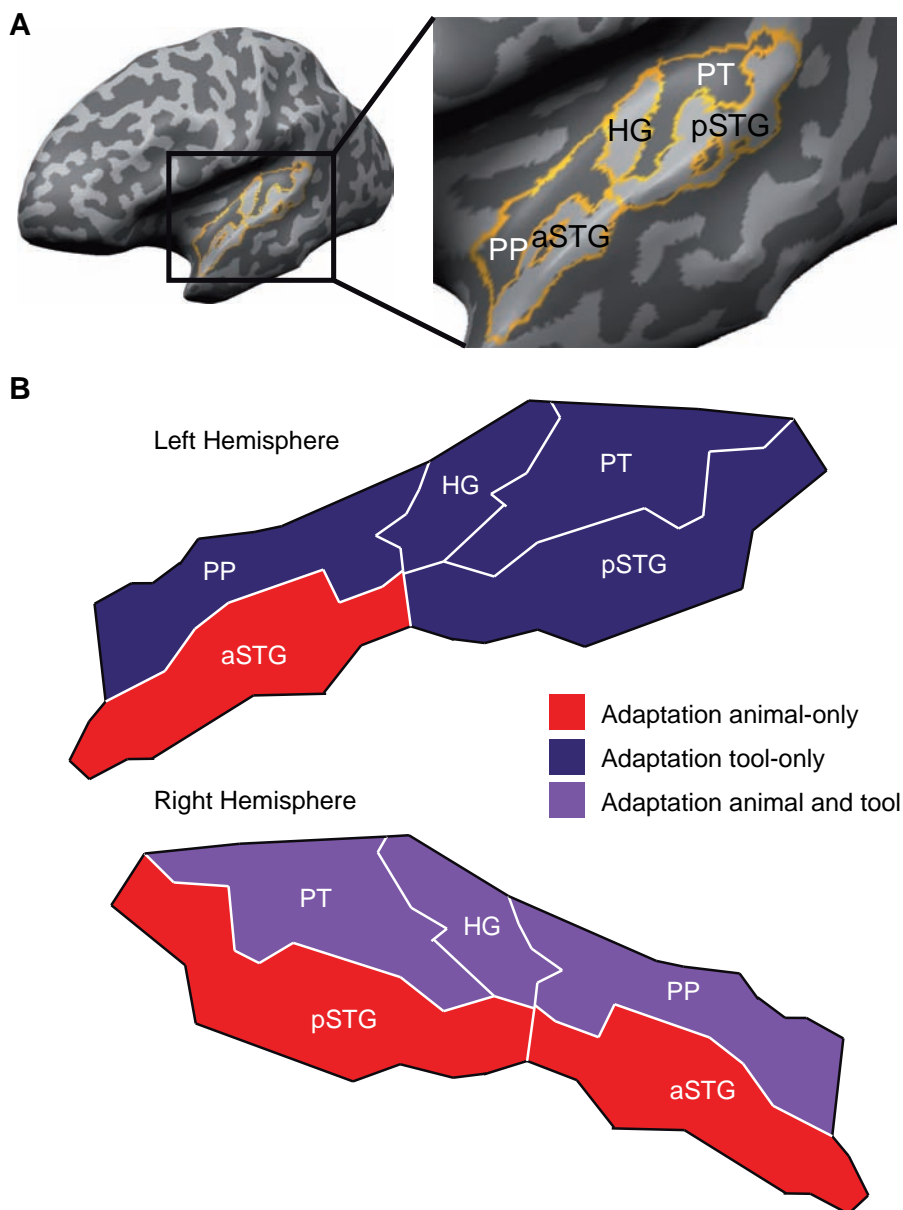


Figure 4. Identification and analysis of anatomical regions of interest. (A) Partially inflated 3D cortex representation of the left hemisphere of one subject. Borders of the anatomical regions of interest on the superior-temporal plane and the superior temporal gyrus as determined by macroscopic criteria are shown in orange. (B) Schematic drawings of the same left (upper panel) and right (lower panel) hemispheric regions. Colors code for significant adaptation effects ( $p < 0.05$ ) for animal vocalizations only (red), tool sounds only (blue) or both sound categories (violet).

to the presentation of pairs of different stimuli compared to pairs of the same stimulus. However, only the anterior STG ROI showed an additional main effect of category ( $F[1,12]=6.752, p<0.05$  [left aSTG];  $F[1,13]=14.302, p<0.05$  [right aSTG]).

These results for aSTG were further supported by adaptation effects for animal vocalizations but not for tool sounds ("selective adaptation"). This is particularly interesting for the left aSTG since all other ROIs in the left hemisphere showed the reverse pattern with significant adaptation effects for tool sounds only (Fig. 4B). In the right hemisphere, both ROIs located on the STG show adaptation effects restricted to animal vocalizations, although the pSTG showed a trend to adaptation to tool sounds as well ( $t(13)=2.076, p<0.058$ ). In all other right-hemispheric ROIs this adaptation effect to animal vocalizations was always accompanied by a corresponding effect for tool sounds.

## Discussion

In the present study, we employed an fMRI-adaptation paradigm to assess category-related preferences and selectivity for animal vocalizations and tool sounds in regions of the human temporal cortex. Extending previously described preferences for animal vocalizations and tool sounds, we were able to show that these effects do translate into corresponding selective adaptation only in STG and left pMTG, respectively. More precisely, a region located anterior to left HG on the STG was found to adapt to animal vocalizations, but not to tool sounds, while the pMTG showed the reverse adaptation profile. However, these regions were not completely complementary with respect to their response patterns: the STG region was still responsive to tool sounds in the sense of a significantly stronger fMRI signal for these sounds than for fixation only, while in the left pMTG animal vocalizations did not produce activation different from baseline. Finally, our anatomically constrained analysis further supports a central role of the anterior STG for the processing of animal vocalizations as it was the only region under investigation with a selective adaptation to this sound category. Interestingly, all other ROIs in the left hemisphere selectively adapted to tool sounds, while all ROIs, except the ones on the STG, in the right hemisphere showed a non-selective adaptation profile.

### Preference for and selective adaptation to animal vocalizations

By contrasting fMRI responses to animal vocalizations with those to tool sounds (main effect "category"), effects were detected along the STG anterior-laterally to HG. Similar results concerning the preference for animal vocalizations in these regions have been reported previously, for instance by Lewis and coworkers (2005). Interestingly, in our earlier study (Altmann et al., 2007) we also found significant responses in bilateral STG by contrasting intact animal vocalizations to degraded control stimuli, which preserved the overall spectral complexity but lost the temporal and the spectro-temporal properties of the original sounds (see similarly Engelien et al., 2006). These findings are in line with a general interpretation of STG as forming part of the non-primary auditory cortex (Kaas and Hackett, 2000) and an intermediate processing stage for complex sounds (Nelken, 2004; Lewis et al., 2005). This notion is furthermore consistent with results showing that auditory cortex beyond the primary core regions is particularly sen-

sitive to complex stimuli such as band-pass filtered sounds in contrast to pure tones (Wessinger et al., 2001) and to amplitude- and frequency-modulated sounds (Hall et al., 2002; Hart et al., 2003).

However, the major aim of the present study was to further our understanding of differential processing for distinct sound categories by using fMRI adaptation. Importantly, selective adaptation effects for animal vocalizations were found in superior temporal regions similarly localized as the ones reported by Lewis et al. (2005) and our previous study (Altmann et al., 2007). Furthermore, these results were confirmed in the context of our anatomically constrained analysis in which regions of the anterior portions of the STG exclusively showed a main effect of category as well as an adaptation effect restricted to animal vocalizations.

Several pieces of evidence from our study support an interpretation of these findings according to the “intermediate processing stage”-hypothesis for the anterior STG mentioned before. First, although STG seems to be particularly sensitive to stimulus properties present in animal vocalizations, it still shows significant responses above baseline for tool sounds also. Thus, on a very general level it is responsive to properties common in both types of sounds, for example a rich spectral content and high modulation rates (Hall et al., 2002).

Second, in left STG the response to the “animal and tool”-condition was found to be significantly stronger when compared to the mean of both “same” conditions, but weaker when compared to the mean of both “different” conditions. The “animal and tool”-condition was introduced to control for transitions from one spectrally complex sound to a second one as well as a change of category. However, the latter aspect did not seem to enhance (or suppress) the response to this control condition beyond the signal for the “different” conditions. Thus, STG seems to track variations in physical stimulus properties rather than changes across category boundaries.

Third, similar to the stimulus analysis by Lewis et al. (2005), we found differences between animal vocalizations and tool sounds with respect to the relative distribution and peak localization of the spectral content (averaged power spectra in Fig. 1) and harmonic-to-noise ratio. Thus, these two sound categories have distinct physical properties, for instance in the spectral and the spectro-temporal domain, which seem to be represented differently in the STG (Zatorre and Belin, 2001; Jamison et al., 2006) as evidenced by the animal-selective adaptation effects.

Finally, a recent event-related potential study (Murray et al., 2006) that similarly investigated processing of animal vocalizations and sounds of man-made objects has demonstrated differential processing already 70 ms after stimulus onset. The calculated source model identified regions of the superior temporal cortex as likely sites for these effects (see also Brugge et al., 2003; Inui et al., 2006). Our finding of the most pronounced repetition-induced adaptation effects in the left STG (for animal vocalizations) suggests a possible role of this cortical region in these differentiation processes probably based on a tight coupling to the physical properties of these sounds. This might then provide an input to higher-order regions implicated in active categorization of acoustic stimuli such as the anterior insula (Binder et al., 2004) or the inferior frontal cortex (Adams and Janata, 2002).

However, further studies utilizing various neuroimaging techniques are certainly needed to further shed light on which sound properties are actually represented diffe-

rentially in superior temporal regions (Griffiths et al., 2004). In our earlier study (Altmann et al., 2007) we have not found adaptation in the STG to the presentation of animal vocalizations or spectrally matched control sounds, suggesting that acoustic features different from the spectral domain might be represented in this cortical region. Alternatively, complex spectro-temporal features particularly drive responses in the STG (see e.g. Boemio et al., 2005). The technique of fMRI-adaptation and especially the use of cross-adaptation, i.e. the change in the fMRI signal caused by the variation of one feature of a complex stimulus, such as the temporal envelope without a corresponding change in sound category, should enable a more fine-grained mapping of auditory cortex and its responsiveness to environmental sounds.

### **Preference for and selective adaptation to tool sounds**

Several cortical regions showed a preference for tool sounds when compared to animal vocalizations. Most clusters were found anterior to HG extending into the PP and further into insular cortex as well as ventral frontal regions. Posterior to HG, voxels with significant responses to tool sounds were detected in the left pMTG and the right STG spreading into the MTG.

However, by examining the activation profiles of these and additional regions with regard to the selectivity of adaptation effects, only regions in the left hemisphere showed selective responses to tool sounds, namely, left PP, left HG/PT and left pMTG. These results were confirmed in our anatomically constrained analysis: all left-lateralized ROIs (except aSTG) selectively adapted to tool sounds only, while all ROIs in the right hemisphere with an adaptation effect to tool sounds (HG, PT, PP) showed a corresponding effect for animal vocalizations as well (i.e. non-selective adaptation).

Importantly, only the left PP and the left pMTG had a response pattern for tool sounds which was comparable to the one found in STG for animal vocalizations. More precisely, these regions were preferentially activated for tool sounds as compared to animal vocalizations and additionally showed a tool-selective adaptation effect. No corresponding effects were found in similar regions of the right hemisphere.

Posterior MTG has been implicated in auditory recognition processes (Maeder et al., 2001; Lewis et al., 2004) which seem to be particularly pronounced for tool sounds (Whatmough et al., 2002; Lewis et al., 2005). Interestingly, several studies have found pMTG to be responsive to tool-related stimuli which could be presented as sounds (Beauchamp et al., 2004; Lewis et al., 2005; 2006), in a linguistic format such as spoken or written words (Noppeney et al., 2005; Noppeney et al., 2006) or as pictures or videos (Martin et al., 1996; Beauchamp et al., 2002). This suggests a multisensory and possibly action-related representation of tools in pMTG (Beauchamp et al., 2004).

Our findings support the notion that pMTG possibly operates on a higher processing level than STG. While the latter is generally responsive to stimuli taken from both auditory categories as evidenced by an fMRI signal stronger than fixation baseline, pMTG is only significantly activated for tool stimuli. Thus, while STG might still be involved in the process of differentiation between the two sound categories, activation of pMTG seems to be more restricted to the category of tool sounds.

Concerning the question why the present study revealed such a striking lateralization to the left hemisphere for tool-related cortical processing, several possible explanations have to be considered. First, both animal vocalizations and tool sounds evoked

stronger responses in the left than in the right hemisphere. This is consistent with findings demonstrating left-lateralized activations for meaningful compared to meaningless (degraded) sounds (Engelien et al., 2006; Altmann et al., 2007). Second, and more specifically, a recent study has found stronger fMRI responses to tool sounds in pMTG of the hemisphere contralateral to the dominant hand (Lewis et al., 2006). Since all subjects in our sample were right-handed, predominantly left-lateralized effects for tool sounds are certainly in line with these results. More generally, it has been suggested (Lewis, 2006; Martin, 2007) that tool-related sensory processing partly overlaps with those regions that are associated with the planning and actual execution of the respective associated body movements (see e.g. Johnson-Frey et al., 2005; Lewis et al., 2005). Thus, overlearned associations between auditory stimulus properties and corresponding motor-related information might have similarly activated our network of temporal regions.

Finally, in the present study tool-selective adaptation was already found in early auditory regions such as HG, PT and PP. The temporal resolution of fMRI is not sufficient to resolve the question if these effects reflect processing of physical stimulus properties pertaining exclusively to tool sounds or rather feedback projections from higher cortical levels. However, the absence of corresponding selective adaptation effects in the same ROIs of the right hemisphere suggests the latter interpretation. Possibly, feedback from regions such as left pMTG might bias processing of tool sounds only in auditory regions of the same hemisphere, while the same region on the right simply adapts to common physical properties for both sound categories. However, this has to be determined by future studies using neuroimaging methods with higher temporal resolution.

### **The impact of conceptual and semantic features of common sounds**

The auditory stimuli used in the present experiment cannot only be characterized on the level of physical properties, but also based on an individual learning history associating physical features with conceptual knowledge related to the objects. In this domain of semantic memory it has been suggested (see e.g. Martin, 2007) that those cortical regions which are particularly involved during the processing of sensory input are also active whenever conceptual representations are retrieved in the course of cognitive tasks and action preparation. The results of the present study are consistent with this proposal since our stimulus set was drawn from the general class of common sounds and was additionally structured by category membership. Accordingly, the aspect that each stimulus was recognizable and meaningful might have contributed to the response pattern reported here. Particularly for pMTG a large number of findings suggest an involvement in semantic-related processing especially of tool sounds. For instance, pMTG has been found to be activated for moving (Beauchamp et al., 2002) and static (Chao, et al., 1999) visual tool stimuli, for both perceived and executed body movements (Johnson-Frey et al., 2005), for visually and auditorily presented stimuli (Beauchamp et al., 2004), and for recognizable sounds of common objects (Lewis et al., 2004) as well as word stimuli (Noppeney et al., 2005). The present finding of a tool-selective adaptation effect together with the lack of an fMRI response significantly differing from silent baseline further strengthens the interpretation of pMTG as an important component within the semantic system, particularly with regard to tool stimuli.

## Conclusion

The present study used an fMRI-adaptation paradigm to investigate category-related cortical responses to two auditory stimulus categories. Preferential responses as well as selective adaptation effects were found particularly in the left STG for animal vocalizations and the left pMTG for tool sounds. However, these regions might not operate on exactly the same processing level since especially STG seems to be more sensitive to auditory stimuli *per se*, while processing in pMTG appears to be restricted to tool sounds.

We therefore suggest that the cortical representation of environmental sounds in the STG indicates a general representation of complex stimulus features, such as their spectro-temporal content (Boemio et al., 2005). Differential processing of one stimulus category, as evidenced by selective adaptation effects to animal vocalizations, can already be found on a cortical level as early as left STG. Left pMTG, however, seems to be less affected by variations in physical stimulus features, suggesting a more specific coding for tool sounds without a corresponding sensitivity to stimuli of the other category. However, future studies using advanced imaging techniques such as fMRI adaptation will certainly deepen our understanding of category-related auditory processing. For instance, extending the results presented here, future experimental paradigms might proceed (rather in a bottom-up manner) to identify critical features of natural sounds likely to be represented in cortical regions such as the STG. This might then permit to match and further minimize the influence of these acoustical features of sounds from different categories. Therefore, the challenge for a complete understanding of how variations in physical stimulus properties give rise to different category-related cortical representations consists in a close matching of as many of these properties as possible while preserving the respective category membership (see Op de Beeck et al., 2008, for a similar argument on visual categories). Using fMRI adaptation, the present study provides a first step towards this aim of identifying cortical representations of particular sound categories.

## Acknowledgements

This study was supported by the Bundesministerium für Bildung und Forschung (Brain Imaging Center Frankfurt, DLR 01GO0203), the Frankfurt Medical Faculty intramural young investigator programme (to M.J.N.) and the Deutsche Forschungsgemeinschaft (DFG AL 1074/2-1). The authors would like to thank Tim Wallenhorst and Maria Katharina Döring for assistance with data acquisition and Sarah Weigelt for helpful comments on earlier versions of the paper.



# 4

## Cortical plasticity of audio-visual object representations

Several regions in human temporal and frontal cortex are known to integrate visual and auditory object features. The processing of audio-visual (AV) associations in these regions has been found to be modulated by object familiarity. The aim of the present study was to explore training-induced plasticity in human cortical AV integration. We used functional magnetic resonance imaging to analyze the neural correlates of AV integration for unfamiliar artificial object sounds and images in naïve subjects (PRE-training) and after a behavioral training session in which subjects acquired associations between some of these sounds and images (POST-training). In the PRE-training session, unfamiliar artificial object sounds and images were mainly integrated in right inferior frontal cortex (IFC). The POST-training results showed extended integration-related IFC activations bilaterally, and a recruitment of additional regions in bilateral superior temporal gyrus/sulcus (STG/STS) and intraparietal sulcus (IPS). Furthermore, training induced differential response patterns to mismatching compared to matching (i.e. associated) artificial AV stimuli that were most pronounced in left IFC. These effects were accompanied by complementary training-induced congruency effects in right posterior middle temporal gyrus (pMTG) and fusiform gyrus (FG). Together, these findings demonstrate that short-term crossmodal association learning was sufficient to induce plastic changes of both AV integration of object stimuli and mechanisms of AV congruency processing.

Nauer, M.J.\*, Doehrmann, O.\*, Müller, N.G., Muckli, L., Kaiser, J., and Hein, G.\* (in press). Cortical plasticity of audio-visual object representations. *Cerebral Cortex*. 10.1093/cercor/bhn200 \* equal contribution



## Introduction

In the last decade, audio-visual (AV) integration of object images and sounds has been investigated in numerous studies. AV integration sites have been identified in superior temporal sulcus (STS; Calvert et al., 2001; Beauchamp et al., 2004; Taylor et al., 2006; Hein et al., 2007), inferior frontal cortex (IFC; Belardinelli et al., 2004; Taylor et al., 2006; Hein et al., 2007), auditory (superior temporal gyrus, STG; Hein et al., 2007; Heschl's gyrus, HG; van Atteveldt et al., 2004; 2007a) and visual cortex (Belardinelli et al., 2004). There is increasing evidence that the identified regions have different AV integration profiles. AV integration in auditory and visual regions has been found mainly for familiar and congruent sounds and images, for example a dog picture and a barking sound (Belardinelli et al., 2004; Hein et al., 2007). STS preferably integrates familiar material, independent of congruency (Beauchamp et al., 2004; Hein et al., 2007). AV integration regions in IFC were preferably involved in the integration of familiar but incongruent material (e.g., dog picture and meowing sound; Taylor et al., 2006; Hein et al., 2007) or unfamiliar stimuli (Hein et al., 2007). Taken together, these findings indicate that the cortical activation patterns related to AV integration are affected by both the familiarity and semantic congruency of the stimuli in both modalities.

In everyday life there are many cases where the familiarity of an object's image and its sound changes, for example when we learn to associate a certain blip or error signal to a new technical device. Based on previous findings showing an impact of familiarity on neural AV integration it is plausible to assume that such training-induced familiarity modifies cortical AV integration. However, the neural plasticity of AV integration has not been investigated to a larger extent. In the present study, we used fMRI to investigate modulation of cortical AV integration after training of crossmodal associations between unfamiliar three-dimensional object images and sounds.

A few previous fMRI studies on crossmodal learning have assessed time-dependent activation changes during AV learning (Gonzalo et al., 2000; Tanabe et al., 2005) or changes in functional connectivity between unimodal cortices (von Kriegstein and Giraud, 2006). Cortical regions exhibiting plasticity correlating with AV learning have been found in the posterior hippocampus, parietal and dorsolateral prefrontal regions (Gonzalo et al., 2000) and the STS (Tanabe et al., 2005). These latter studies employed low-level AV stimulus sets including musical chords or white noise for auditory and Chinese characters or two-dimensional abstract patterns for visual stimulation. In contrast to the two-dimensional AV stimuli used by Gonzalo et al. and Tanabe et al., the unfamiliar object images used in our study had a three-dimensional shape together with shadings etc., resembling everyday life cases (e.g. images of unknown technical devices). Moreover, three-dimensionality of our unfamiliar objects makes them also more comparable to familiar AV pairings, which in most cases are also comprised of a three-dimensional object paired with a complex sound. So far, training-induced changes in AV integration for such object-like, initially non-meaningful stimuli have not been investigated yet.

There is interesting evidence for training-induced neural changes in object-related sound processing (Amedi et al., 2007) as well as processing of threedimensional,

initially unfamiliar, object images (James and Gauthier, 2003; Weisberg et al., 2007). Amedi et al. compared neural activations in blind and sighted subjects after they had learnt to recognize everyday objects based on complex artificial sound patterns. Training-induced changes of activation elicited by these sounds were found in intraparietal sulcus (IPS) and prefrontal cortex, mainly along the pre-central sulcus. After training, blind subjects revealed sound-induced activations in the lateral occipital tactile-visual area (LOtv), providing striking evidence for a crossmodal representation of 3D shape in this region (Amedi et al., 2007).

In one previous visual study subjects learnt linguistic labels ("names") for pictures of complex artificial stimuli (so-called "greebles"), with these labels referring either to movements or sounds (James and Gauthier, 2003). After learning, the purely visual presentation of these artificial stimuli led to activations of the respective motion- and sound-processing cortical regions. In another study subjects were presented with artificial "tool-like" objects and trained on how to use these novel tools (Weisberg et al., 2007). Before training, the presentation of photographs of these objects only led to ventral visual activations, whereas the neural responses after training resembled those previously found for photographs of highly familiar common tools.

It is an open question whether there is similar neural plasticity in AV integration of crossmodally trained objects. To investigate this question, our participants were presented with artificial 3D object images (so-called "fribbles") and sounds (Fig. 1A) in a pre-training fMRI experiment (PRE), a crossmodal training session, and a subsequent second fMRI experiment (POST). In the training session, subjects learnt novel associations between the sounds and images, thus increasing their familiarity with these artificial stimuli. In a control condition, we presented images and sounds of highly familiar animals. The results of the PRE training session have been reported elsewhere (Hein et al., 2007).

Our first prediction was that the cortical integration pattern for artificial AV pairings would change with training. We were recently able to demonstrate that familiar AV pairings are integrated in temporal regions such as the STG and STS, whereas integration of unfamiliar artificial object sounds and images mainly activated right IFC (Hein et al., 2007). Based on these findings, we expected that a training-induced increase of familiarity should correlate with an increase of integration-related AV activation in frontal and additional temporal regions. Our second prediction was that the establishment of associations between artificial sounds and images should also be reflected by effects of congruency on cortical activation patterns. We expected activation differences between conditions with artificial AV pairings for which an association was acquired and those comprised of randomly paired artificial sounds and images. There are two different sets of regions, which have been found to show an AV activation profile sensitive to the semantic congruency between object images and sounds. One set of regions comprises higher-order auditory or visual regions, which are known to be involved in the integration of congruent sounds and images (Belardinelli et al., 2004; Hein et al., 2007). Accordingly, training-induced congruency effects should be found in such temporal regions. The other region is the IFC, which was found to integrate unfamiliar artificial AV pairings (Hein et al., 2007), and, moreover, showed stronger AV activations for incongruent as compared to

congruent familiar AV combinations (Taylor et al., 2006; Hein et al., 2007). Based on the latter, we also expected IFC activation in the POST training session in response to artificial AV combinations which, after training, were considered as mismatching.

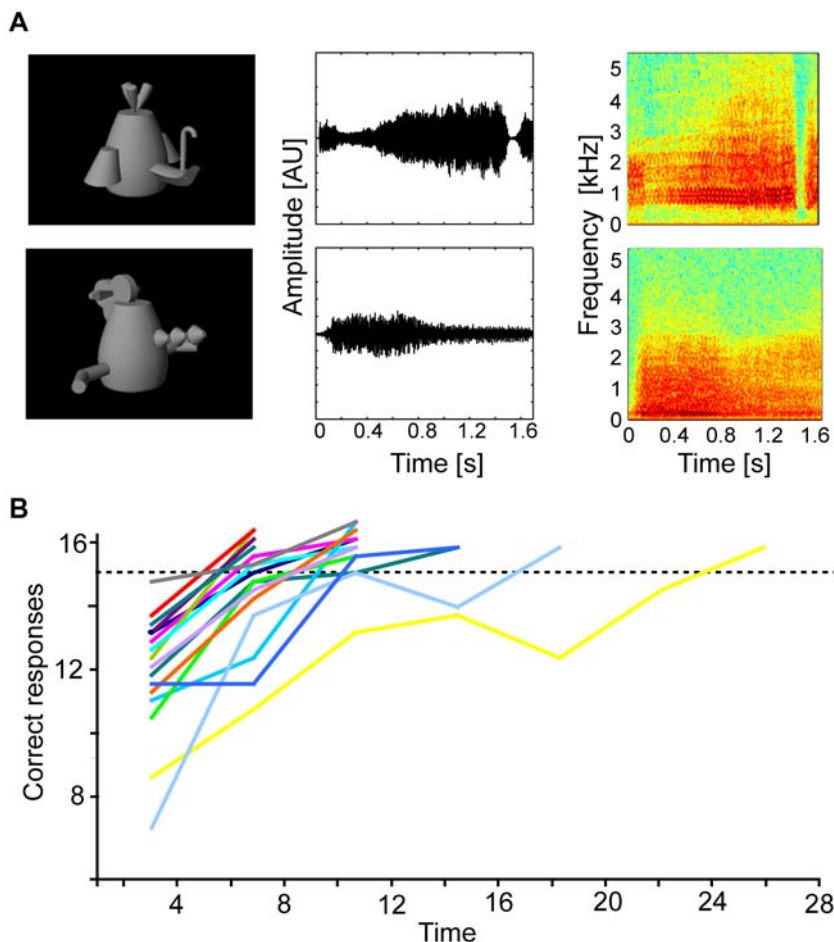


Figure 1. Audio-visual (AV) stimuli and behavioral data. (A) Audio-visual (AV) stimuli consisted of combinations of gray-scale photographs (left column) and complex sounds of both natural (i.e. animals) and artificial material (i.e. so-called ‘fribbles’). AV combinations could be either congruent, i.e. corresponding to the learnt associations between the artificial visual and auditory stimuli (upper row in A) or incongruent (lower row in A). Middle and right columns provide the amplitude waveforms and spectrograms for typical exemplars of artificial sounds. (B) Behavioral training data of 17 subjects and inter-individual differences in AV learning speed. The dashed horizontal line indicates the criterion of 15 (out of a maximum of 16) correct responses which had to be reached in four successive training blocks in order to terminate the recording session. AU, arbitrary units.

## Materials and Methods

### Overall structure of the study

The study consisted of three major parts. All subjects underwent a first fMRI session (hereafter referred to as PRE), which served as a baseline for the remaining experiment. Subsequently the same subjects participated in a behavioral training session, which aimed at establishing an association between artificial auditory and visual stimuli. On the day following the training session, the second fMRI scan (POST) was performed with the same experimental setup as in the PRE session forming the third and final part of the study. Prior to the POST scan, subjects underwent behavioral testing to ensure that the learning criterion (15 out of 16 correctly identified AV combinations; see below) was still fulfilled. The delay between the PRE and the POST sessions varied between 74 and 123 days (mean: 94.9 days).

### Participants

Eighteen adults (seven female, mean age = 29.8 years, range 23–41 years, one lefthanded), with normal or corrected-to-normal vision and hearing participated in the study. They received information on MRI and a questionnaire to check for potential health risks and contraindications. Subjects gave their written informed consent in accordance with the declaration of Helsinki.

### Stimuli

During the fMRI sessions, gray-scale 3D images were presented in the center of a black screen with a mean stimulus size of 14.6 degrees visual angle. A fixation cross was constantly present. Artificial images were chosen from the “fribble” database (Fig. 1A; <http://titan.cog.brown.edu:8080/TarrLab/stimuli>). Sounds were presented via headphones simultaneously to both ears. They consisted either of animal vocalizations or artificial sounds, created through distortion (played backwards and filtered with an underwater effect) of the same animal vocalizations. In a behavioral pre-test, distorted sounds were presented to eight subjects. None of the distorted sounds was associated with any common object. Sounds and images were presented in stimulation blocks at a rate of one per 2 s. Stimulation blocks consisted of eight stimulus events.

### Procedure for the fMRI experiments (PRE and POST)

The fMRI sessions were comprised of eight different experimental conditions: unimodal conditions for animal vocalizations (A-NAT), artificial (i.e. distorted) sounds (A-ART), animal images (V-NAT), and artificial fribble images (V-ART). Bimodal conditions consisted of synchronously presented sounds and images with varying degrees of semantic congruency (AV-NAT-CON: e.g., dog photograph – barking sound; AV-NAT-INCON: e.g., dog photograph – meowing sound; AV-ART-CON, and AV-ART-INCON). Note that in the first fMRI session (PRE), i.e., prior to the establishment of learnt associations between the artificial sounds and images, no differences between AV-ART-CON and AV-ART-INCON were expected. For the PRE

session, the sole difference between these two experimental conditions was that the AV stimuli were presented either in fixed or in randomized pairings.

The eight conditions were presented within a block design with 16s of stimulation (eight recording volumes) alternating with 16s of central fixation. A complete experimental run consisted of two cycles of all experimental conditions with additional eight volumes of fixation at the beginning of the run. We had five experimental runs, with the order of blocks being pseudo-randomized in each run and counterbalanced across runs. In the fMRI sessions, subjects were instructed to fixate and attentively perceive the experimental stimuli.

A passive paradigm was chosen to minimize task-related activations of the frontal cortex (see also Calvert et al., 2000; Belardinelli et al., 2004; van Atteveldt et al., 2004; 2007a). Given the sluggishness of the BOLD fMRI signal, it might otherwise be hard to disentangle a potential involvement of frontal regions in AV integration from the task-related frontal activations.

### **Audio-visual association training**

Prior to the experiments, eight AV associations – each consisting of one gray-scale fribble image and one distorted sound – were declared as “correct.” Training consisted of several runs, each containing four blocks with 16 trials that were presented in randomized order to reduce predictability. Eight of these trials comprised the “correct” associations; in the other eight trials the artificial visual and auditory stimuli were paired randomly. Subjects had to categorize the AV pairings as “correct” or “incorrect” via button press after each trial. Since they performed at a self-paced speed, only accuracy data were analyzed. Immediately after pressing the button, a visual feedback was given on the screen (“The answer was correct.” or “The answer was incorrect.”). After each block the number of correct trials was displayed as an additional feedback. Subjects were instructed to avoid learning strategies based on verbalization, for example by using invented “names” for the stimuli. The number of runs was not fixed since the aim was to ensure that each subject received the necessary amount of training in order to successfully learn the “correct” associations. The criterion was reached when in four successive blocks within one run at least 93.75 % (i.e. 15/16) of the answers were correct. If a subject did not meet this criterion, another run of four blocks started. At the beginning of this new set of blocks as well as at the very beginning of the whole behavioral training session the “correct” associations were presented to further facilitate the training process.

### **MRI data acquisition**

Functional MRI was performed with a 3 Tesla Siemens Magnetom Allegra scanner (Siemens, Erlangen, Germany). A gradient-recalled echo-planar-imaging (EPI) sequence was used with the following parameters: 34 slices; TR, 2000 ms; TE, 30 ms; FoV, 192 mm; in-plane resolution, 3 x 3 mm<sup>2</sup>; slice thickness, 3 mm; gap thickness, 0.3 mm. For each subject, a magnetization-prepared rapid-acquisition gradient-echo (MP-RAGE) sequence was used (TR = 2300 ms, TE = 3.49 ms, FA = 12°, matrix = 256 x 256, voxel size 1.0 x 1.0 x 1.0 mm<sup>3</sup>) for detailed anatomical imaging.

## Whole-brain data analysis

Data were analyzed using the BrainVoyager™ QX software package (version 1.8; Brain Innovation, Maastricht, The Netherlands). The first four volumes of each experimental run were discarded to preclude T1 saturation effects. Preprocessing of the functional data included the following steps: (i) three-dimensional motion correction, (ii) linear-trend removal and temporal high-pass filtering at 0.0054 Hz, (iii) slice-scan-time correction with sinc interpolation, and (iv) spatial smoothing using a Gaussian kernel of 8 mm full-width at half-maximum (FWHM).

Volume-based statistical analyses were performed using random-effects general linear models (RFX-GLM;  $df = 17$ ). For every voxel the time course was regressed on a set of dummy-coded predictors representing the experimental conditions. To account for the shape and delay of the hemodynamic response (Boynton et al., 1996), the predictor time courses (box-car functions) were convolved with a gamma function. RFX-GLMs were computed both separately for each experimental session and across sessions. The latter model permitted explicit testing for effects across time.

Neural correlates of object-related AV integration were assessed for artificial material in the PRE and POST fMRI sessions. We searched for regions that were a) significantly activated during each of the unimodal conditions (A; V), and b) responded more strongly to bimodal AV stimulation than to each of its unimodal components. Accordingly, the identification of brain regions involved in object-related AV integration was based on significant activation in a  $0 < A < AV > V > 0$  conjunction analysis, a statistical criterion well-established in previous studies (Beauchamp et al., 2004; van Atteveldt et al., 2004; 2007a; Hein et al., 2007).

In addition, we directly tested for PRE-POST effects on the basis of the whole-brain random-effects general linear model (RFX GLM) across both data sets. The respective conjunction analysis required both significantly stronger responses to bimodal artificial AV conditions compared to both unimodal conditions (A-ART and V-ART) as well as stronger responses to bimodal artificial conditions during the POST versus the PRE session ( $AV-ART-POST > PRE$ ).

In a second step of analysis we tested for effects of congruent and incongruent AV stimulation. We reasoned that training might have induced such effects for artificial AV material in cortical regions which are also sensitive to congruency relations in natural AV stimuli. Accordingly, we searched for voxels exhibiting stronger signals for the contrast  $AV-ART-CON > INCON$  (or vice versa) in those regions which showed additionally significant results in the contrast  $AV-NAT-CON > INCON$  (or vice versa). Statistically this corresponds to a main effect of *congruency* for both types of stimuli. To explicitly test this effect, we entered the beta values of the four AV conditions from the RFX-GLM of the POST fMRI-session into a (whole-brain) two-factorial analysis of variance (ANOVA) as implemented in the respective Brainvoyager QX tool. Factors (and levels) were *stimulus type* (NAT, ART) and *congruency* (CON, INCON). The main effect of *congruency* was then specified in this model by contrasting CON versus INCON pooling across the factor *stimulus type*. Voxels identified by these contrasts were taken as regions-of-interest (ROIs) and further analyzed with respect to the question which response patterns they exhibited in the first experimental session (see below). This aimed at further

supporting potential training-induced effects for artificial AV stimulation in the POST session.

In addition to these contrasts, we also searched in unimodally defined sensory cortices for training-induced effects of both AV integration and congruency. To this end, we contrasted responses to each of the natural unimodal conditions (A-NAT; VNAT) separately with responses to the respective artificial counterpart (A-ART; VART). We chose these contrasts to identify those unimodal regions which are likely to be sensitive to semantic features of object stimuli, or, more precisely, regions corresponding to higher-order auditory cortices and the ventral visual cortex. Voxels determined by these contrasts were similar to the previously described regions further analyzed in a ROI analysis.

Generally, statistical maps were corrected for multiple comparisons using cluster-size thresholding (Forman et al., 1995; Goebel et al., 2006). In this method, for each statistical map the uncorrected voxel-level threshold was set at  $t > 4$  ( $p < 0.001$ , uncorrected; unless otherwise indicated) and was then submitted to a whole-brain correction criterion based on the estimate of the map's spatial smoothness and on an iterative procedure (Monte Carlo simulation) for estimating cluster-level false-positive rates. After 1,000 iterations the minimum cluster-size that yielded a cluster-level false-positive rate of 5% was used to threshold the statistical map. When possible, group-averaged functional data were projected on inflated representations of the left and right cerebral hemispheres of one subject. As a morphed surface always possesses a link to the folded reference mesh, functional data can be shown at the correct location of folded as well as inflated representations. This link was also used to keep geometric distortions to a minimum during inflation through inclusion of a morphing force that keeps the distances between vertices and the area of each triangle of the morphed surface as close as possible to the respective values of the folded reference mesh.

### ROI-based analysis

Regions identified in the previous whole-brain analyses were further investigated in the course of a ROI-based analysis with respect to the question of congruency effects induced by novel AV associations. In addition to the voxels identified by the main effect *congruency* in the data of the POST fMRI session we defined further ROIs based on the respective unimodal (i.e. auditory and visual) contrasts between natural and artificial conditions. We conducted RFX GLMs and computed t-tests in these ROIs comparing responses to congruent and incongruent AV stimulation for the natural and the artificial conditions in both sessions. We were mainly interested in regions exhibiting a stable effect for the comparison of congruent and incongruent natural stimuli in both experimental sessions and, additionally, a novel, i.e. training-induced effect of congruency for artificial AV stimuli only in the POST session. Statistically, this would be most robustly reflected by a transition from a significant interaction of the factors *material* and *congruency* in the PRE session (due to congruency or incongruency effects only for the natural material) to a main effect of *congruency* in the POST session. Additionally, two-factorial ANOVAs with the factors *time* (PRE, POST) and *congruency* (CON, INCON) were computed separately for the natural and the artificial AV stimuli to verify that changes occurred only for the latter. Furthermore, we entered the ROI beta weights of the GLM



including data from both experimental sessions into a three-factorial ANOVA with the factors *material* (NAT, ART), *congruency* (CON, INCON), and *time* (PRE, POST). We thereby explicitly tested for any significant interactions of these three factors which should most likely reflect changes in the congruency relation of artificial stimuli across time.

In all figures containing bar plots of ROI-based beta estimates mean standard errors are additionally reported. The latter were corrected for inter-subject effects similar to previous studies (Altmann et al., 2007; Doebrmann et al., 2008). In particular, the fMRI responses were calculated individually for each subject by subtracting the mean percent signal change for all conditions within this subject from the mean percent signal change for each condition and adding the mean percent signal change for all the conditions across subjects.

### Evaluation of repetition-related effects

In a final analysis step we aimed at determining to which extent effects for artificial AV stimuli might be affected by unspecific repetition-related effects, for instance due to adaptation. To this end, we computed in the across-session GLM the contrast AV-ART-CON (POST) > AV-ART-INCON (PRE) and vice versa. The rationale for this was that the stimuli of the experimental condition AV-ART-CON were repeated most frequently since they were presented in this constellation in both fMRI sessions as well as in the training session. Thus, any repetition-related effects were most likely present in fMRI signals of this experimental condition. On the other hand, the condition AV-ART-INCON (PRE) is least likely to contain any repetition-related effects due to 1.) the randomization of the relation between artificial stimuli and 2.) the fact that the data were recorded in the first experimental session.

To further corroborate the robustness of our other results, we increased the likelihood of finding any unspecific repetition-related effects by using uncorrected thresholds of  $t(17) = 4$  and  $t(17) = 3$  for these contrast maps. These maps were then projected to inflated representations of the cerebral cortex of one of our subjects. We then also projected the main effect of incongruency onto these cortical surfaces to evaluate the potential overlap of these effects with the former ones. Furthermore, we additionally computed ROI-based paired t-tests for AV-ART-CON (POST) > AV-ART-INCON (PRE) to exclude unspecific repetition effects for the trained material in the regions exhibiting our most relevant experimental results. And finally, we also contrasted AV-NAT-CON (POST) > AV-NAT-CON (PRE) in order to control for stimulus-specific repetition effects regarding the untrained, but still repeated natural material.

## Results

### Behavioral training results

In the training period, subjects were presented with combinations of artificial object sounds and images until they reached a criterion of 15 out of 16 correct responses. The mean number of training blocks necessary to reach this criterion was 12.7 (range: 8-28) which was approximately equivalent to three runs (consisting of four blocks each) per subject. After the first run the mean number of correct responses



was 11.63 (of 16 possible). At the end of the training session (i.e., after reaching the criterion) the number of correct trials increased to 15.25 per block in the final run. The general trend of increasingly successful associations between artificial AV stimuli was well-established for all subjects (Fig. 1B). Since prior to each training run (consisting of four blocks each) the correct AV associations were presented to the subjects, the results for the first blocks were already well above chance level. The same held true for the elevated performance after each run, which was probably due to retrieval from short-term buffers, rather than an establishment of long-term associations between the stimuli. Our criterion for a stable performance of 15/16 correct responses on average over four successive blocks was chosen to account for this.

On the day of the POST fMRI session, we checked if subjects still met the criterion. For this purpose, the same procedure as in the training session was repeated. Most of the subjects (i.e., 15/18) could immediately replicate their performance from the day before. Only three of them needed another run of four training blocks to reach the criterion again. The mean number of blocks needed for refreshment was 4.89 with a mean performance of 15.41 correct responses per block. This shows that for each subject the trained artificial AV associations were still present at the time of the POST fMRI session.

### **fMRI results: Whole-brain data analysis**

To investigate whether the training of crossmodal associations between formerly unfamiliar artificial object images and sounds had an impact on AV integration-related cortical activations, we analyzed the data of the PRE and POST fMRI sessions separately. For this purpose, we employed the so-called MAX-criterion (i.e.  $AV > \max[A, V]$ ).

#### *PRE-training*

The results of the PRE fMRI session have already been reported in detail in our previous publication (see Hein et al. 2007). As can be seen there, unfamiliar artificial images and sounds were mainly found to be integrated in right IFC, whereas familiar natural AV objects were integrated in a bilateral fronto-temporal network of regions including IFC, STS, and STG. (Fig. 2A). Note that the results of the PRE session are shown with a slightly higher initial statistical threshold ( $t(17) = 3.25$ ) than in the previous report of the data (Hein et al., 2007;  $t(17) = 3$ ) before correction for multiple comparisons. As expected, the respective unimodal contrasts A-ART > V-ART (yellow) and the reverse contrast (light blue) revealed robust activations in visual and auditory cortex.

#### *POST-training*

We tested for potential plasticity of cortical AV integration as defined by the max-criterion (i.e.  $AV > \max[A, V]$ ) for artificial AV pairings (pooled across congruent and incongruent conditions) in the POST fMRI session. The main results of this analysis ( $t(17) = 3.25$ ,  $p < 0.05$ , corrected, cluster size threshold =  $189 \text{ mm}^3$ ) are depicted in Fig. 2B (dark blue-colored patches) while the respective Talairach coordinates and cluster sizes of these regions are provided in Table 1A. The AV integration-related

cortical network included bilateral IFC and STG/STS regions, in line with our first prediction. Moreover, AV integration-related activity was found in bilateral anterior insula and IPL, as well as precuneus and VOT in the left hemisphere.

In addition, we directly tested for PRE-POST effects on the basis of a whole-brain

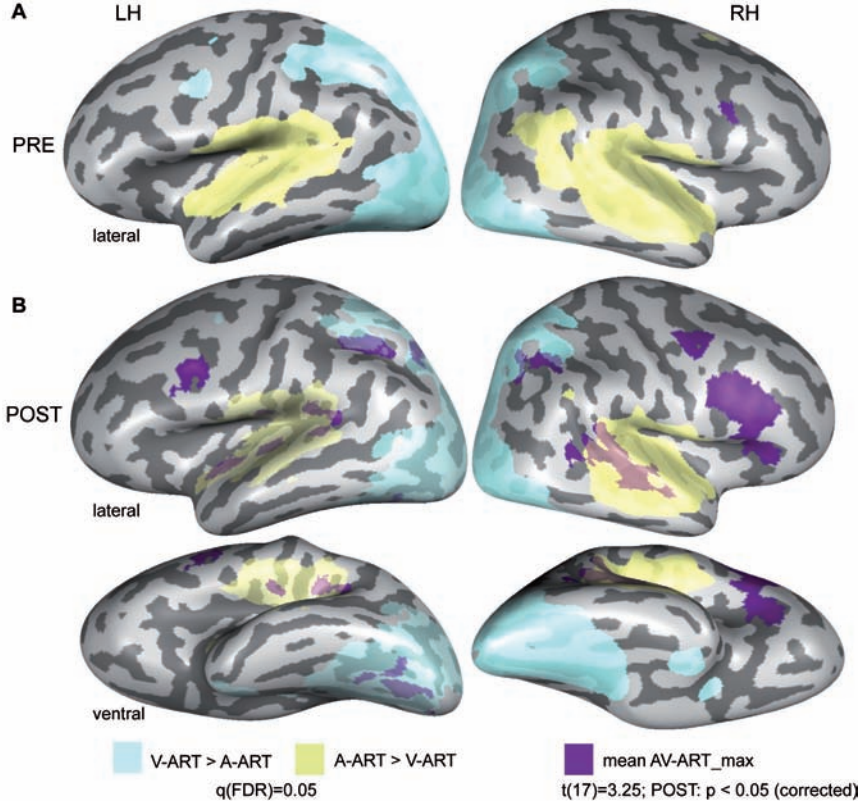


Figure 2. Cortical regions involved in AV integration of artificial stimuli during the first (A) and the second (B) experimental session were determined by a conjunction analysis comparing the mean BOLD fMRI responses to both (congruent and incongruent) artificial AV conditions to each unimodal condition (MAX-criterion:  $AV > \max[A-ART, V-ART]$ ). Similar to our previous study (Hein et al., 2007) the additional criteria  $A-ART > 0$  and  $V-ART > 0$  were added to this conjunction. Maps were then adjusted to an initial t-value of 3.25 and then submitted to a volume-based cluster-threshold algorithm yielding a new map thresholded at  $p < 0.05$  (corrected; cluster-size 189 voxels). These corrected maps were then projected onto three-dimensional reconstructions of both cortical hemispheres of one subject from our sample. See Table 1 A for Talairach coordinates and extents of cortical clusters depicted here. Furthermore, unimodal activations as determined by the contrasts  $A-ART > V-ART$  (auditory; yellow) and  $V-ART > A-ART$  (visual; light blue) were computed and corrected based on the false-discovery rate (FDR) yielding a significance value of  $q(FDR) < 0.05$ . LH and RH denote the left and right hemispheres, respectively. PRE and POST refer to the first and second experimental session.

Table 1. Talairach coordinates of regions showing significant effects (A) for the max-contrast (POST) and (B) for the PRE-POST comparison.

Cortical region	x	y	z	Cluster size (voxels)
<i>A. Max contrast</i>				
Left superior temporal gyrus (posterior)	-46	-36	13	799
Left superior temporal gyrus (middle)	-44	-22	7	329
Left superior temporal gyrus (anterior)	-41	-10	-4	344
Left ventral occipito-temporal cortex	-35	-60	-17	755
Left inferior frontal gyrus	-46	5	26	464
Left inferior parietal lobule	-31	-50	37	815
Left precuneus	-23	-66	35	341
Left posterior cingulate	0	-28	23	487
Right posterior superior temporal gyrus	52	-31	9	2742
Right middle frontal gyrus (ventral)	44	14	21	2696
Right middle frontal gyrus (dorsal)	42	-2	48	523
Right inferior parietal lobule (anterior)	31	-52	39	372
Right precuneus	28	-62	34	198
Right posterior cingulate	1	-28	23	556
<i>B. PRE-POST comparison</i>				
Left middle frontal gyrus (ventral)	-41	21	22	522
Left inferior frontal gyrus	-45	5	26	465
Left Insula	-29	21	4	205
Left inferior parietal lobule	-31	-50	38	399
Left posterior cingulate	0	-29	21	374
Right inferior frontal gyrus	47	15	21	1344
Right precentral gyrus	49	4	31	319
Right Insula	33	21	3	554
Right inferior parietal lobule	34	-52	38	355
Right presupplementary motor area	7	9	50	222
Right posterior cingulate	1	-28	22	410

Clusters were determined by comparing the mean of the two artificial AV conditions to the respective unimodal controls (volume-based cluster threshold of the map: 189 voxels; surface-based threshold 25 mm<sup>2</sup>). Cluster sizes for the cortical regions reported in Fig. 2B (part A of the table) and in Fig. 3 (part B) are provided in a resolution of 1x1x1 mm<sup>3</sup> voxels.

RFX GLM across both fMRI data sets. The respective conjunction analysis included the following three criteria: 1) AV-ART-POST>A-ART-POST, 2) AV-ART-POST>V-ART-

POST, and 3) AV-ART-POST>AV-ART-PRE. The main results of this analysis ( $t(17) = 3.25$ ,  $p < 0.05$ , corrected, cluster size threshold =  $189 \text{ mm}^3$ ) are depicted in Fig. 3 while the respective Talairach coordinates and cluster sizes of these regions are provided in Table 1B. The results revealed activations in bilateral posterior cingulate regions and fronto-parietal (IPL, IFC), but not temporal portions of the aforementioned AV integration-related network.

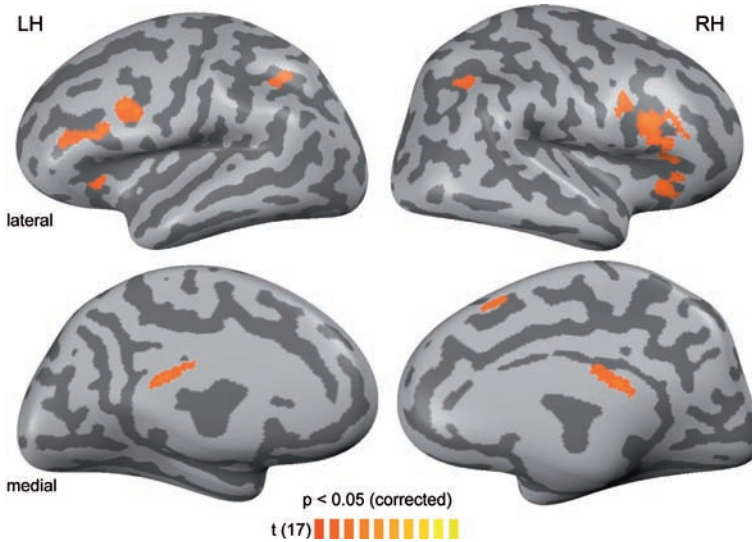


Figure 3. Results of a contrast investigating effects from the first (PRE) to the second (POST) fMRI session on the basis of beta weights from a whole-brain random-effects general linear model (RFX GLM) across both data sets. A conjunction analysis required both significantly stronger responses to bimodal artificial AV conditions compared to both unimodal conditions (A-ART and V-ART) as well as stronger responses to bimodal artificial conditions during the POST versus the PRE session (AV-ART-POST > PRE). Maps were adjusted to an initial t-value of 3.25 and a volume-based cluster size threshold of 189 voxels (see Fig. 2), resulting in a new map thresholded at  $p < 0.05$  (corrected). These corrected maps were then projected onto inflated reconstructions of both cortical hemispheres of the same subject as in Fig. 2. See Table 1B for Talairach coordinates and extents of cortical clusters depicted here.

In our second prediction we hypothesized that the training of crossmodal associations between formerly unrelated sounds and images might also induce effects of congruency, i.e., activation differences in two possible directions: AV-ART-CON > INCON or AV-ART-INCON > CON. To test whether our training procedure had established congruency effects for our artificial AV stimuli corresponding to those for familiar natural stimuli, we analyzed the POST session data by means of a whole-brain ANOVA with the factors *material* (NAT, ART) and *congruency* (CON, INCON). This analysis revealed a statistical main effect ( $t(17) = 4$ ,  $p < 0.05$ , corrected) for incongruency in left anterior IFC (Fig. 4A; see Table 2 for Talairach coordinates). A further region of interest (ROI)-

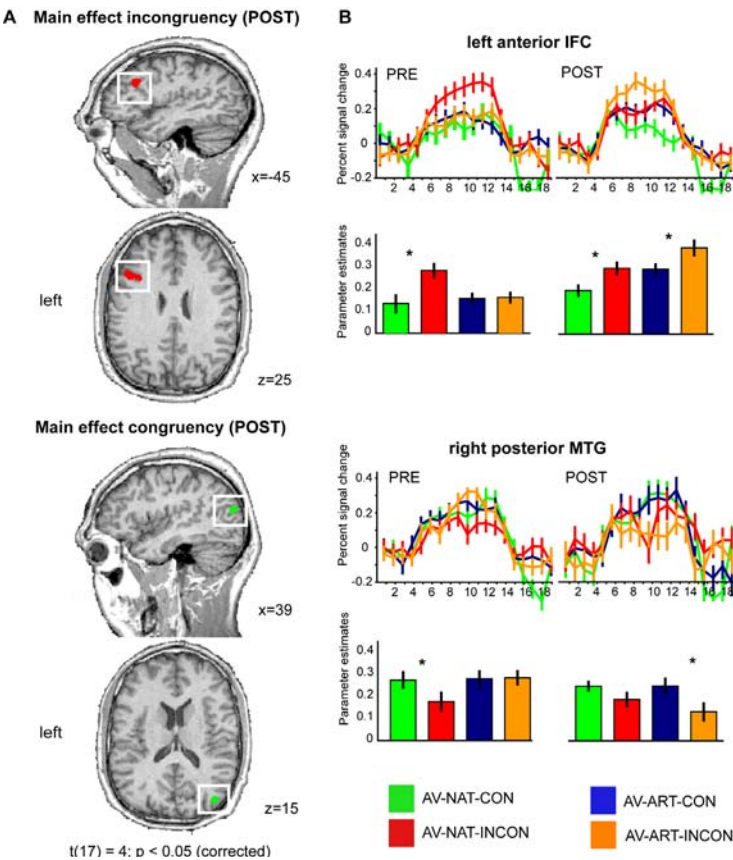


Figure 4. POST-training main effects of incongruency and congruency. Beta weights from the whole-brain RFX GLM of the POST-training session were entered into a two-factorial analysis of variance (ANOVA) with the factors *stimulus type* (NAT, ART) and *congruency* (CON, INCON). In this model we specified a contrast corresponding to the main effect of congruency since the latter optimally captured the assumed development of congruency and incongruency for artificial AV material in those cortical regions which were similarly sensitive to congruency in natural AV stimuli. (A) Results of the contrasts are shown in the upper panel in the direction of incongruency (INCON>CON) and in the lower for congruency (CON>INCON). Maps were adjusted to an initial t-value of 4 and then submitted to a volume-based cluster-threshold algorithm yielding a new map thresholded at  $p < 0.05$  (corrected; cluster-size 115 voxels). Maps were then projected on sagittal and transversal slices of an anatomical dataset of the same subject which provided the cortex reconstructions in Fig. 2. (B) Region-of-interest (ROI)-based analyses of cortical regions depicted in A). Significant clusters in the left anterior inferior frontal cortex (IFC; upper panel) and right posterior middle temporal gyrus (pMTG; lower panel) were further analyzed with regard to the question of congruency or incongruency effects. For each ROI the respective contrasts (e.g. AV-NAT-INCON > AV-NAT-CON) were calculated in separate RFX GLMs for both types of stimulus material and for both sessions. Time course information and beta estimate plots are provided for these two regions. Error bars represent mean standard errors. Asterisks denote significant effects on a level of  $p < 0.05$ . Color coding is provided at the bottom of the figure. Additional statistical information can be found in Table 2.

based analysis of the activation profiles (by means of BOLD-signal time courses and beta estimates; Fig. 4B) demonstrated that incongruency effects for familiar natural stimulus material were stable across the PRE and POST fMRI sessions. In contrast, similar incongruency effects for the artificial AV objects were only observable during the POST fMRI session, i.e. after successful crossmodal association training. This was also evident in a significant *time x congruency* interaction ( $F(1,17) = 5.042$ ,  $p < 0.05$ ) for the artificial material only. Additionally, these training-induced shifts in left IFC were also indicated by the transition from a significant statistical interaction of the factors *material* and *congruency* in the PRE-session (driven by the response to incongruency in familiar natural AV stimuli) to a main effect of incongruency which was constitutive for this ROI. A summary of these analyses is given in Table 2.

Table 2. Summary of ROI analysis on congruency and incongruency effects.

Cortical region / Session	Statistical test			
	AV-NAT-INCON > CON	AV-ART-INCON > CON	Material x Congruency	Material x Time x Congruency
Left IFC (-38, 13, 29)				
POST	t(17)=3.063 <b>p=0.007042</b>	t(17)=3.244 <b>p=0.004771</b>	t(17)=0.098 p=0.923029	F(1,17)=11.310 <b>p=0.004</b>
PRE	t(17)=3.021 <b>p=0.007710</b>	t(17)=0.235 p=0.817001	t(17)=2.929 <b>p=0.009363</b>	
	AV-NAT-CON > INCON	AV-ART-CON > INCON	Material x Congruency	Material x Time x Congruency
right pMTG (40, -78, 18)				
POST	t(17)=2.027 p=0.058627	t(17)=4.201 <b>p=0.000601</b>	t(17)=1.851 p=0.081568	F(1,17)=4.086 p=0.059
PRE	t(17)=2.328 <b>p=0.032499</b>	t(17)=0.178 p=0.860894	t(17)=1.718 p=0.103925	
right FG (35, -40, -16)				
POST	t(17)=-0.322 p=0.751495	t(17)=2.340 <b>p=0.031751</b>	t(17)=-1.622 p=0.123256	F(1,17)=0.101 p=0.754
PRE	t(17)=-0.793 p=0.438727	t(17)=0.628 p=0.538534	t(17)=-0.849 p=0.407857	

Clusters were determined by computing main effects of AV congruency and incongruency in the two-factorial ANOVA on the factors Material (NAT, ART) and Congruency (CON, INCON). Volume-based cluster threshold of the map was 115 voxels; surface-based threshold was 25 mm<sup>2</sup>. Coordinates refer to Talairach space (Talairach and Tournoux, 1988). Abbreviations: IFC: inferior frontal cortex; pMTG: posterior middle temporal gyrus; FG: fusiform gyrus.

The reverse contrast to this examination of training-induced incongruency, i.e. a main effect of congruency, revealed a cluster in right posterior MTG. The same ROI analysis as for the left IFC demonstrated that a stronger response to congruent as compared to incongruent artificial AV conditions was only present in the POST fMRI session. This was again supported by a significant *time x congruency* interaction ( $F(1,17) = 9.764$ ,  $p <$



0.05). However, a corresponding congruency effect for familiar natural stimuli was only significant in the PRE fMRI session and just failed to reach significance in the POST session ( $t(17)=2.027$ ,  $p=0.059$ ). Therefore, also the *material x congruency* interaction was not significant in this cluster ( $t(17)=1.851$ ,  $p=0.082$ ; see Table 2 for a complete summary).

Additionally, we conducted a three-factorial ANOVA across the data of both sessions with the factors *material* (NAT, ART), *time* (PRE, POST), and *congruency* (CON, INCON). This analysis revealed a significant *material x time x congruency* interaction in left aIFC which was largely driven by the change in the fMRI response to the AV-ART-INCON condition. In line with our above reported analysis the corresponding effect in pMTG is less pronounced as revealed by a trend towards significance of this interaction (see Table 2 for the respective F- and p-values).

In line with our prediction this analysis showed that left anterior IFC was more strongly activated in response to formerly unrelated artificial sounds and images. After training these IFC responses were more pronounced for stimulation with incongruent than congruent pairs of artificial AV stimuli, while right posterior MTG showed a complementary, but less pronounced congruency preference pattern. These effects resemble the congruency effects for familiar natural AV objects which have been found consistently across both sessions in these regions.

### Visual and auditory ROIs

Based on previous results (Belardinelli et al., 2004; van Atteveldt et al., 2004; Hein et al., 2007), we also predicted congruency effects in higher-level visual and auditory regions. To define higher-level visual and auditory regions involved in the semantic processing of sounds and images, we contrasted the respective natural and artificial conditions (i.e. A-NAT > A-ART and V-NAT > V-ART, respectively; see Materials and Methods). This analysis revealed significant activations in bilateral STG and right fusiform gyrus (FG), which were then used as regions of interests (ROIs; Fig. 5). In the right FG (Fig. 5A), the crossmodal association training induced a significant congruency effect, i.e., stronger AV activation to congruent as compared to incongruent artificial material (PRE:  $t(17)=0.793$ ,  $p=0.44$ ; POST:  $t(17)=2.340$ ,  $p<0.05$ ), without corresponding congruency effects for natural stimuli (see Table 2 for details). Table 2 again summarizes the outcome of these ROI analyses.

In contrast, unimodal auditory regions in STG showed no significant AV congruency effect for artificial material (Fig. 5B). However, right STG revealed a training-induced AV integration effect for artificial material. The latter was characterized by a significantly stronger response to trained artificial AV stimuli than to each of the unimodal (i.e. auditory and visual) conditions in the POST session (PRE:  $t(17)=1.400$ ,  $p=0.17$ ; POST:  $t(17)=2.551$ ,  $p<0.05$ ).

### Evaluation of repetition related effects

One alternative explanation for the stronger POST training activity in IFC, IPL, and temporal regions could be based on the potential presence of unspecific repetition effects. This would reflect the fact that subjects have been stimulated with the artificial material in the PRE training, the behavioral training, and the POST training sessions. In order to control for these potential repetition-related effects we contrasted the experimen-

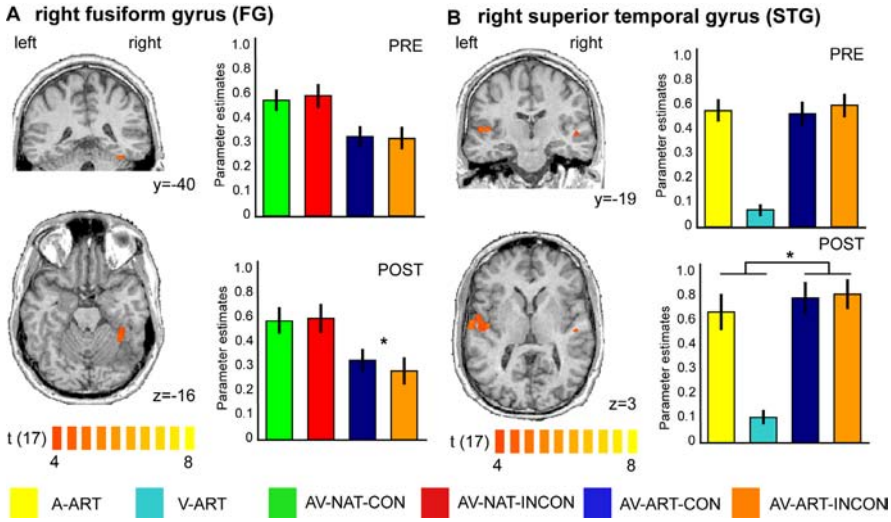


Figure 5. Unimodally defined ROIs: training-induced congruency and multisensory integration effects. Significant voxels are depicted for unimodal visual (A) and auditory (B) contrasts directly comparing stimulations with natural objects to those with artificial material in the POST fMRI session (i.e. V-NAT > V-ART (POST) and A-NAT > A-ART (POST), respectively). Maps were adjusted to an initial t-value of 4 and then submitted to a volume-based cluster-threshold algorithm yielding a new map thresholded at  $p < 0.05$  (corrected) for the visual (minimum cluster threshold: 271 voxels) and the auditory (minimum cluster threshold: 136 voxels) contrasts, respectively. The panels on the left show coronal (upper rows in A and B) and transversal slices (lower rows in A and B) of the same anatomical dataset used in Fig. 4. The same ROI-based analysis as described in Fig. 4 was applied to the right fusiform gyrus (FG; A) for all bimodal conditions and for both fMRI sessions. A similar ROI-based analysis was applied to the region in right superior temporal gyrus (STG, B), revealing a training-induced AV integration effect ( $0 < A < AV < V > 0$ ) for artificial material. Error bars of the beta estimate plots represent mean standard errors. Asterisks denote significant effects on a level of  $p < 0.05$ . Color coding is provided at the bottom of the figure. Additional statistical information for the FG ROI can be found in Table 2.

tal condition which was arguably most likely to be affected by such effects (AV-ART-CON [POST]) with the experimental condition least likely to be affected in this regard (AV-ART-INCON [PRE]; see Materials and Methods).

Significant effects ( $t(17) = 4$ ;  $p < 0.001$ , uncorrected) for the contrast AV-ART-CON [POST] > AV-ART-INCON [PRE] are shown in Fig. 6A (orange colors). Particularly regions of the medial cortex located in the bilateral posterior cingulate and pre-supplementary motor area and the right precuneus were detected by this contrast. Only small clusters were revealed in the inferior-most and anterior frontal cortex. Most importantly, none of these clusters was overlapping with our main effect of incongruency reported in the previous analysis (white circle in Fig. 6A). We chose an uncorrected threshold to maximize the likelihood of detecting such overlaps. However, even when further lowering the threshold for this map to  $t(17) = 3$  no overlap with either the IFC (incongruency) or the right pMTG (congruency) could be detected (data not shown). Even ROI-based



testing for repetition effects in these regions did not reveal any significant effects (for details see Table 3).

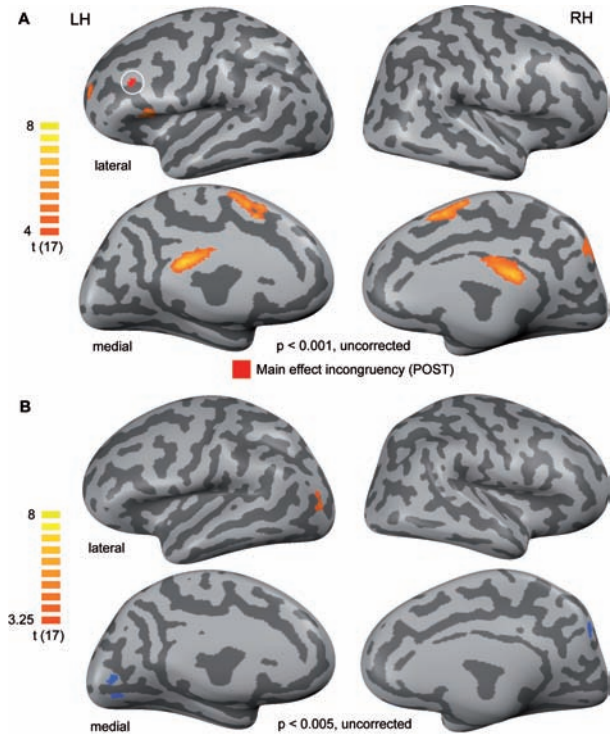


Figure 6. (A) To control for repetition-related effects we contrasted the experimental condition which is arguably most likely to be affected by such effects (AV-ART-CON [POST]) with the experimental condition least likely to be affected by this (AV-ART-INCON [PRE]). Only positive effects (signal increases from PRE to POST) were significant at the chosen threshold. These are shown in an orange-to-yellow color code. Maps are thresholded to a t-value of 4 (uncorrected for multiple comparisons) to increase the sensitivity for potential repetition-related effects and were then projected onto the same cortex reconstructions used in Fig. 2 shown in lateral (upper panels) and medial (lower panels) views. Only a minimal surface-based cluster threshold of 25 mm<sup>2</sup> was applied. Additionally, the main effect of incongruity as described in Fig. 4 is shown (in red and highlighted by a white circle) to evaluate a potential frontal overlap of these maps. LH and RH denote the left and right hemispheres, respectively. (B) To control for stimulus-specific repetition effects we additionally contrasted the unlearned, but also repeated experimental condition comprised of natural AV stimuli. The corresponding contrast can be stated as follows: AV-NAT-CON (POST) > AV-NAT-CON (PRE). Since no effects were present on a statistical t-value of 4, we lowered the threshold to an uncorrected t-value of 3.25. The same anatomical basis as in the previous figures was used for the projection of results.

Finally, when contrasting AV-NAT-CON-POST versus PRE (Fig. 6B) in order to control for stimulus-specific repetition effects, significant activations ( $t(17) = 3.25$ ;  $p < 0.005$ , uncorrected) were only found in a left lateral occipital region (POST>PRE; orange-colored region in Fig. 6B) and bilateral medial occipital regions (POST<PRE; blue-colored regions). As none of these regions overlapped with any of our main ROIs, these results

render it unlikely that the reported training-induced congruency and integration effects were due to stimulus repetition.

Table 3. Summary of ROI-based analysis of the control contrasts.

Cortical region	Statistical test	
	<i>AV-ART-CON (POST)</i> <i>&gt; AV-ART-INCON (PRE)</i>	<i>AV-ART-CON (POST)</i> <i>&gt; AV-ART-CON (PRE)</i>
Left IFC	$t(17)=-1.83$ $p=0.085$	$t(17)=-1.53$ $p=0.145$
right pMTG	$t(17)=1.333$ $p=0.2$	$t(17)=1.378$ $p=0.186$
right FG	$t(17)=-0.124$ $p=0.903$	$t(17)=0.257$ $p=0.8$
right STG	$t(17)=-0.293$ $p=0.773$	$t(17)=-0.982$ $p=0.340$

Statistics were based on two-tailed paired t-tests comparing the most repeated experimental condition (AV-ART-CON [POST]) to the two artificial AV conditions in the first experimental session. Repetition-related effects are most likely revealed by these contrasts. Abbreviations: IFC: inferior frontal cortex; pMTG: posterior middle temporal gyrus; FG: fusiform gyrus; STG: superior temporal gyrus.

## Discussion

In the present study we investigated training-induced cortical plasticity of object-related AV integration. We compared the neural correlates of AV integration for unfamiliar artificial material before and after a behavioral training session, in which subjects acquired crossmodal associations between object images and sounds. We assumed that our training session might modify neural AV processing in two different ways. First, we expected that the training-induced increase of familiarity might lead to an increase in integration-related AV activations. Second, we predicted that our training might also induce certain congruency effects as reflected in POST training activation differences between trained artificial AV pairings and those which were paired randomly.

For the trained artificial AV stimuli, the POST-training results showed extended integration-related IFC activations bilaterally, and a recruitment of additional parietal (PPC) and temporal (right STG/STS, and left FG) cortical regions. Furthermore, training-induced effects of congruency were found for these artificial AV stimuli in several cortical regions. We revealed a training-induced AV incongruency effect in left aIFC and a complementary congruency effect in a posterior portion of right MTG. Both cortical regions showed similar effects for familiar natural AV stimuli in each of the two fMRI sessions.

### Integration-related AV activations and incongruency effects

We found training-induced changes of AV integration-related activation patterns for (initially unfamiliar) artificial object stimuli, indicating cortical plasticity of object-re-

lated AV integration. Confirming our first prediction, crossmodal association training resulted in enhanced integration-related AV activation in a fronto-parieto-temporal network including cortical regions in bilateral IFC, IPL, STG/STS, and left FG. Within this network frontal regions showed the most robust training-induced AV plasticity effects, reflected by significant POST>PRE effects. So far, training-induced changes in cortical processing of complex object stimuli had only been shown in the visual modality (James and Gauthier, 2003; Weisberg et al., 2007).

Training-induced activation increases in brain regions which have been activated already before training, such as right IFC in our study, are a well-known result from studies investigating both AV association learning (Gonzalo et al., 2000) and training-induced effects in other domains, for example, perceptual processing (Schwartz et al., 2002), memory (Nyberg et al., 2003; Olesen et al., 2004) or learning of complex 3D object-related sounds (Amedi et al., 2007). Across studies investigating learning processes in diverse contexts, such effects were found in similar inferior frontal regions as those activated in our study, often in concert with parietal regions (Nyberg et al. 2003; Olesen et al., 2004). It is important to note that in contrast to studies focusing on learning, our study investigated training-induced changes in AV integration, not time-dependent neural modulations caused by learning per se. It is nevertheless possible that the observed inferior frontal as well as intra-parietal activation increases might reflect consolidation and storage processes with regard to recently acquired crossmodal associations (see D'Esposito, 2007 for a review).

Alternatively, it could be argued that increased inferior frontal and intra-parietal activations reflect increased attention to artificial AV pairings after crossmodal association training (e.g. Duncan and Owen, 2000; Corbetta and Shulman, 2002). There is solid evidence that novelty is one of the main triggers for an increase of attention-related activation (e.g. Herrmann and Knight, 2001). Based on that, one would predict stronger fronto-parietal activity in the PRE session, where the artificial material was completely novel and unfamiliar. In contrast to this prediction, our results showed an extension of activation in the POST training session, rendering the assumption unlikely that our findings merely reflect attention-related effects.

Another possible interpretation of our fronto-parietal findings could be that the AV integration-related cortical plasticity found here and the learning effects reported for similar regions in classic learning studies might reflect two complementary sides of the same coin. As predicted in our second hypothesis, the AV integration region in left aIFC was sensitive to (acquired) incongruency of artificial AV stimuli in the POST fMRI session, resembling the results for semantically incongruent natural object features in the PRE fMRI sessions. Taken together, these two findings indicate that AV integration in aIFC could serve the establishment and revision of AV associations, i.e., a function, which might also be part of the time-dependent activation changes in IFC found in the learning literature. In contrast, AV integration for artificial images and sounds in superior temporal cortex (STG/STS) was only found in the POST fMRI session, i.e., when the formerly unfamiliar AV material has already been trained. However, these effects were not accompanied by AV-ART-POST>PRE effects. Moreover, STG/STS showed no sensitivity to the degree of semantic AV congruency, neither in the whole brain congruency analysis (Fig. 4), nor when mapped based on the unimodal auditory (A-NAT vs. A-ART) contrast (Fig. 5B). This insensitivity to AV congruency suggests that STG/STS

is rather involved in the integration of sounds and images, which are already familiar to a certain extent, than the establishment of novel AV associations. In this sense, IFC and STG/STS might have complementary roles in AV integration. While STG/STS is integrating AV pairings that already reached a certain degree of familiarity, for example through training as in the present study, the IFC is either revising already known (but apparently incongruent) AV associations or establishing new ones. The explicit testing of this assumption would require information about the temporal flow of information. While the sluggishness of the BOLD fMRI signal certainly forms a general limitation in this regard, fMRI-based measures of effective connectivity such as Granger causality mapping (GCM) might be fruitfully applied in future studies.

In our current study, we investigated training-induced plasticity of AV integration using a passive paradigm which has been inspired by previous fMRI studies on AV integration (Calvert et al., 2000; Belardinelli et al., 2004; van Atteveldt et al., 2004; 2007a). However, the potential impact of task demands on training-induced AV plasticity - especially in the aforementioned fronto-parietal regions - definitely forms another important issue to be investigated in future studies.

### **Training-induced congruency effects**

Besides incongruency effects in left anterior IFC, we found training-induced congruency effects for the POST fMRI session in the posterior aspect of right MTG and in right FG.

Our finding of an AV congruency effect in right pMTG is in line with the general involvement of posterior temporal regions in semantic processing. Significant activations of similar regions have been repeatedly shown particularly in the context of action- and tool-related processing (see e.g. Beauchamp et al., 2002; Johnson-Frey et al., 2005), not only with visual, but also with auditory presentations of the respective stimuli (see e.g. Lewis et al., 2005; Doehrmann et al., 2008). Interestingly, Murray et al. (2004) reported effects for repeatedly presented visual objects which had been paired previously with semantically congruent sounds. Source modeling suggested right-hemispheric effects in the proximity to the cluster reported here and close to lateral occipital regions of higher-level visual cortex. Therefore, a critical role of pMTG in the context of conceptual processing has been suggested previously (see e.g. Martin, 2007; Doehrmann and Naumer, 2008). Our findings show that pMTG is also involved in cross-modal processing of objects which are more artificial than the ones investigated in previous studies (Beauchamp et al., 2002; Johnson-Frey et al., 2005) and should inspire further explorations of this region in future studies.

Ventral regions such as the FG have been repeatedly associated not only with perceptual processing of meaningful object stimuli, but also with the activation of conceptual representations in the course of thought and action (Martin and Chao, 2001; Martin 2007; Doehrmann and Naumer, 2008). Additionally, ventral temporal activation has also been found for AV stimulation with natural stimuli (Beauchamp et al., 2004; Amedi et al., 2005), and some degree of crossmodal plasticity in these regions has recently been demonstrated for higher-level stimuli such as faces and voices (von Kriegstein and Giraud, 2006). Although our finding of stronger responses to congruent, i.e. learned, AV combinations is interesting from this perspective, we refrain from stronger interpretations of this finding in FG since, in contrast to our pMTG cluster, it was not

accompanied by corresponding congruency effects for highly familiar natural stimuli. Certainly, further studies are needed to elucidate the functional role of these ventral temporal regions both for object-related AV processing in general and with respect to crossmodal plasticity in this domain.

Unlike in visual cortex, a unimodally defined auditory region in STG did not show any effect of semantic congruency for artificial AV stimuli. Interestingly, this region exhibited stronger responses to bimodal artificial conditions (compared to the respective unimodal conditions) only in the POST, but not in the PRE (training) session, but without showing any significant POST>PRE effect when directly comparing the respective artificial AV conditions. These findings are inconsistent to some extent with the results from our PRE fMRI session (Hein et al., 2007), which revealed stronger activation for semantically congruent than incongruent natural (i.e. animal) stimuli in right STG. One possibility is that this inconsistency of results is driven by general differences between natural and artificial stimuli. Belardinelli et al. (2004) also reported a congruency effect in higher-level visual but not auditory regions, using some animal stimuli, but also images and sounds of tools. The latter might be more similar to our trained artificial AV objects, which might explain why a training-induced AV congruency effect was found in a similar region as reported by Belardinelli et al. (2004), instead of STG. Another possible explanation is that our training session was too short to induce a sufficient degree of familiarity. From the perspective of the results reported in this and our previous study (Hein et al., 2007) it is likely that prolonged and intensified training might not only enhance responses to bimodal compared to unimodal stimulation, but could also lead to the development of more pronounced congruency effects also in superior temporal regions. The extent of additional training with novel AV stimuli needed for the emergence of congruency effects similar to those for natural objects (Hein et al., 2007) or linguistic stimuli (van Atteveldt et al., 2004) remains to be determined.

In conclusion, our study demonstrates that even short-term crossmodal training of novel AV associations results in integration-related cortical plasticity and training-induced congruency effects for artificial AV stimuli in cortical regions especially of the frontal and (to a lesser degree) the temporal lobes, adding novel aspects to the understanding of object-related AV integration in the human brain.

## Acknowledgements

This work was supported by the German Ministry of Education and Research (BMBF), Frankfurt Medical School (Intramural Young Investigator Program to M.J.N.), the Max Planck Society (to L.M.), the Deutsche Forschungsgemeinschaft (HE 4566/1-1 to G.H.), and Society in Science / The Branco Weiss Foundation (to G.H.). We are grateful to Petra Janson for help with preparing the visual stimuli and Tim Wallenhorst for data acquisition support.

# 5

## Multisensory integration investigated by fMRI adaptation

Information integration across different sensory modalities contributes to object recognition, the generation of associations and long-term memory representations. However, cortical representations and integrative mechanisms especially of meaningful object stimuli have remained largely elusive. We employed event-related functional magnetic resonance imaging adaptation (fMRIa), which offers complementary results compared to conventional neuroimaging paradigms, to further shed light on these issues relevant for the processing of common audio-visual (AV) objects.

Stimulation consisted of an adapting stimulus S1 and a subsequent stimulus S2 in four experimental conditions: 1. the same image and vocalization for S1 and S2, 2. the same image and a different vocalization, 3. different images and the same vocalization, or 4. different images and vocalizations. S1 and S2 were always taken from the same basic-level category (e.g. cat). Together, these conditions formed a two-by-two factorial design which allowed us to assess the contributions of auditory and visual stimulus repetitions in a statistically orthogonal manner.

During the fMRIa experiment responses in visual regions of right fusiform gyrus (FG) and right lateral occipital (LO) were reduced for repeated visual stimuli (repetition suppression; RS). Surprisingly, left LO showed stronger responses to repeated auditory stimuli (repetition enhancement, RE). Similarly, auditory ROIs of the right medial superior temporal gyrus (STG) exhibited RS to auditory repetitions and RE to visual repetitions.

These results point towards a complex interplay of human object-related cortices for repeated AV stimuli as evidenced by the presence of both RS and RE and underscore the view that innovative experimental protocols such as fMRIa can indeed reveal integrative mechanisms complementary to conventional neuroimaging techniques.

Doehrmann, O., Altmann, C.F., Weigelt, S., Kaiser, J., and Naumer, M.J. (in preparation). Multisensory integration investigated by fMRI adaptation.

## Introduction

The ability to process and integrate information obtained from different sensory channels is highly relevant for a plethora of different cognitive capabilities. These could include object recognition, the successful formation of associations between different stimuli and, subsequently, the generation of novel long-term memory representations. In non-human primates single-cell recordings demonstrated an involvement particularly of neurons in temporal and frontal regions during the formation of memory representations derived and audio-visual stimuli (Gibson and Maunsell, 1997; Fuster et al., 2000).

An interesting effect initially revealed by these recordings (Miller et al., 1991; Li et al., 1993) is the finding of reduced or suppressed neuronal firing rates after repeated presentations of the same stimulus. This “repetition suppression” has been regarded as the neuronal correlate of behavioral enhancements in terms of increased processing speed and accuracy due to prior exposure with repeated stimuli (e.g. Desimone, 1996; Wiggs and Martin, 1998). Interestingly, the repeated presentation of stimuli induces similarly a decrease in signal strength of the blood oxygen level-dependent (BOLD) response as measured by functional magnetic resonance imaging (fMRI). In the case of two immediately successive stimuli, the suppressed BOLD signal induced by stimulus repetitions compared to the signal after the variation of a feature in the second stimulus produces in many cases highly selective activation peaks in those regions of the brain which are indeed sensitive to that varied feature. This method, called fMRI adaptation (Grill-Spector and Malach, 2001) developed into a very successful tool of cognitive neuroscience with many different variants commonly in use (see Weigelt et al., 2008 for a review). As for the BOLD signal in general, the exact neuronal underpinnings of the macroscopic fMRI-adaptation effects are still a matter of debate (Krekelberg et al., 2006; Grill-Spector et al., 2006). However, its potential to reveal complementary functional information about cortical regions by detecting selectivity even on the level of neuronal subpopulations is widely acknowledged.

Although fMRI adaptation has been fruitfully applied in the context of the visual (see e.g. Kourtzi et al., 2003; Mahon et al., 2007; Weigelt et al., 2007a) and the auditory sensory modalities (Bergerbest et al., 2004; Altmann et al., 2007; Doehrmann et al., 2008), only a few studies in the multisensory domain have taken advantage of its capacities (for exceptions see e.g. Buckner et al., 2000; James et al., 2002). This is particularly surprising from the perspective of the increased interest in the investigation of brain processes in more natural settings (see e.g. Spiers and Maguire, 2007) and from the general viewpoint that large portions of sensory cortex, including even primary areas, are involved in processing and integrating information from more than one sensory modalities (Ghazanfar and Schroeder, 2006; Driver and Noesselt, 2008). In the case of audio-visual (AV) stimuli, more recent research started to unravel the mechanisms underlying integrative processes after stimulation with different sets of stimuli such as AV speech (Skipper et al., 2007), voice-face integration (e.g. von Kriegstein and Giraud, 2006), letter-speech integration (van Atteveldt et al., 2004; 2007) and object recognition (Beauchamp et al., 2004; Hein et al., 2007; Amedi et al., 2007). Given this large set of studies and results obtained by conventional neuroimaging techniques, the demand



for the application of more advanced methods such as fMRI adaptation has increased accordingly.

Earlier studies investigating repetition-induced processing changes in the AV domain (Buckner et al., 2000; Badgaiyan et al., 2001) focused on priming effects for cognitive tasks such as word-stem completion or word generation. For auditory as well as visual cues, repetition-related signal reductions were particularly found in regions of the left inferior frontal and the ventral temporal cortices. However, these studies did not investigate the impact of repeated AV stimuli in a more direct manner, since stimulation was always confined either to the auditory or the visual sensory modality. Furthermore, the use of a demanding cognitive task probably engaged rather higher-order cortical regions instead of those related to more perceptual processes.

More recent studies (Noppeney et al., 2008; Galati et al., 2008) examined the impact of visual prime stimuli on subsequently presented environmental sounds. Both studies additionally varied the semantic relation (congruent versus incongruent) between the first and the second stimulus. Galati and colleagues (2008) found enhanced responses in left inferior frontal and middle temporal gyrus (MTG) for action sounds which were preceded by a semantically congruent compared to an incongruent prime word. In contrast, Noppeney et al. (2008) revealed effects only for incongruent semantic relations between visual prime stimuli (printed words or pictures) and target sounds (spoken words or sounds) in the left inferior parietal and the medial frontal gyrus and in the left MTG close to superior temporal sulcus (STS).

A common feature of most of these previous studies was the use of verbal material as a first stimulus (prime). However, since the content of word stimuli had to be decoded and possibly transformed into an ("amodal") format to relate it to a subsequent target stimulus, these paradigms most likely favored higher-order processes and corresponding regions of parietal and frontal cortices. At least to the best of our knowledge, no study yet has employed an fMRI-adaptation paradigm which operated on a perceptual level and utilized only non-linguistic, but still meaningful AV stimuli. Stimuli of the latter kind are, for instance, common objects including animals. Images and vocalizations of animals are highly familiar even when presented in isolation. Apart from a putative major importance from an evolutionary viewpoint (New et al., 2007), the auditory and visual component of an AV animal stimulus are associated on a semantic level. This semantic association, in turn, has profound impact on the efficiency of AV integration (van Atteveldt et al., 2004; Hein et al., 2007; Doehrmann and Naumer, 2008).

Thus, in contrast to previous studies, the present work focused on an earlier level of processing of semantically "loaded" AV stimuli minimizing potential confounds from linguistic factors and by reducing influences from (in-)congruency. Using animal images and vocalizations as a model for common objects in general, we compared fMRI responses to presentations of either 1.) the same image and vocalization in the first and second AV stimulus, 2.) the same image and different vocalizations, 3.) different images and the same vocalization, or 4.) different images and vocalizations (Fig. 1A). The change between the two AV stimuli always occurred within the same basic-level category, for instance "frog" or "dog" to keep the degree of semantic "relatedness" constant within each AV stimulus and also across the stimulus pairs. The experimental design consisted of a two-by-two factorial manipulation with the factors "auditory" (same, different) and "visual" (same, different). We were therefore able to investigate



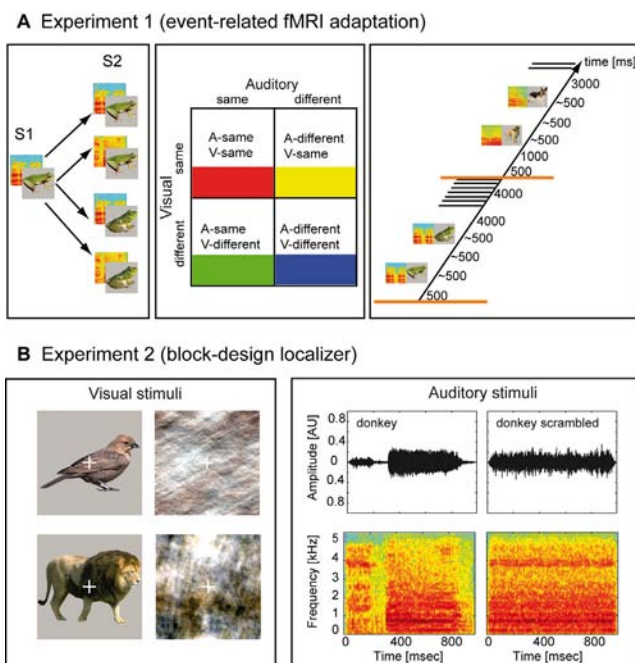


Figure 1. Stimuli and experimental designs. (A) Experiment 1. Schematic representations of the four different trial types employed in the experiment are shown in the left panel. From the first to the second AV stimulus the auditory, the visual or both components changed or no change at all occurred. Note that all changes occurred within the same basic-level category, thereby minimizing the potential influence of semantic congruency. Together, this resulted in a two-factorial design with the factors “auditory” and “visual” (middle panel). Each experimental trial (right panel) started in half of the cases with a jitter of one second followed by the rapidly successive presentation of a pair of two AV stimuli. After three or four seconds two volumes of functional data were collected. (B) Experiment 2. Examples of intact and scrambled visual (left panel) and auditory stimuli (right panel) are given. The latter are represented by amplitude waveforms and spectrograms. Note that the presentation time of auditory stimuli was one second instead of 500 ms as in Experiment 1. Blocks of ten either intact or scrambled stimuli were presented in the auditory, the visual or both sensory modalities, thereby generating six different experimental conditions.

in a completely orthogonal manner the contribution of visual and auditory stimulus repetitions for the processing in the respective sensory cortices.

Previous work from our (Hein et al., 2007) and other groups (e.g. Beauchamp et al., 2004) implicated particularly regions in lateral temporal as well as inferior frontal cortices, and to a lesser extent also ventral temporal regions (Bellardinelli et al., 2004) in the processing of AV object stimuli. Generally, since we employed complex object stimuli, we expected to find differential responses in higher-order auditory and visual cortices. To further enhance the likelihood of detecting such effects we took advantage of separate localizer scans which were designed to identify object-sensitive cortical regions of the superior temporal plane and the lateral occipital (LO) as well as ventral temporal cortices, respectively.

We expected to find effects of suppressed BOLD responses for the repetition of the auditory component in the anterior superior temporal gyrus (STG; Altmann et al., 2007; Doehrmann et al., 2008) and for the visual component in LO as well as fusiform gyrus (FG; Grill-Spector and Malach, 2001; Vuilleumier et al., 2002). However, fMRI-adaptation paradigms with visual (e.g. Henson et al., 2000; Turk-Browne et al., 2007) and auditory stimuli (Gagnepain et al., 2008) demonstrated not only effects of repetition suppression (RS), but also signal increases for repeated stimuli ("repetition enhancement"; RE). More precisely, RE has been found particularly for the repetition of unfamiliar stimuli, and this effect was also present on the level of oscillatory activity as measured by EEG (Fiebach et al., 2005; Gruber and Müller, 2005). Thus, no strong a priori hypothesis was formulated if effects were more likely to occur in the direction of RS or RE.

Visual repetition effects in cortical regions commonly associated with auditory processing or auditory repetition effects in visual cortical regions would directly offer evidence for content-sensitive integration across sensory modalities in these cortical regions. Our experimental paradigm was designed to demonstrate exactly this kind of effects on cortical levels as early as STG and LO which are commonly involved in object-related integrative processing.

## Methods

### Subjects

Fifteen healthy volunteers with a mean age of 27.13 years (range 18-39; 7 males) participated in two experiments. Three subjects (2 males) were left-handed as determined by self-report and an abbreviated form of the Edinburgh Handedness Inventory. Visual acuity for all subjects was normal or corrected to normal. According to self-report hearing abilities were also in a normal range. Subjects gave their informed consent to participate in the study. The experiments were performed in accordance with the ethical standards laid down in the 1964 declaration of Helsinki and were approved by the local Ethics Committee of the Frankfurt Medical School.

### Stimuli

For the fMRI-adaptation experiment (Experiment 1) two different color images and sounds of the following 14 animals were created: bird, cat, cow, dog, donkey, duck, elephant, frog, horse, lion, monkey, owl, rooster, and sheep. Thus, both exemplars of the images and sounds referred to the same basic-level category. Generally, great care was taken to select only those images and sounds from bigger databases which were quite common and easily recognizable even when separately presented. This was especially important because during the experiments images and sounds were presented in different combinations. All animal images were taken from a professional database (Hemera Technologies, Gatineau, Quebec, Canada). The background of each image was removed and replaced by light gray (example stimuli can be found in Fig. 1). Furthermore, a white fixation cross was added. When presented on the screen in the scanner bore, all stimuli (including background) had a size of 10.3 x 10.3 cm (22 x 22 degrees of visual angle).

Twenty-eight animal vocalizations (two for each of the animals mentioned before) were taken from a collection of professionally recorded sounds (Sound Ideas, Richmond Hill, Ontario, Canada) and from different internet websites. The sound stimuli were digitized with a sampling rate of 22050 Hz. The sounds were trimmed to an average length of 487.5 ms (standard deviation 46 ms, range: 342-556 ms). As in the case of the pictures, an important aspect was to preserve the recognizability of each sound. Accordingly, and likewise for the images, no subject reported problems to recognize any of the sound stimuli used during this experiment.

For the localizer experiment (Experiment 2) 18 different animal stimuli were selected from the same sources as for Experiment 1. The stimulus set here was partly overlapping with the one from Experiment 1. One visual and one auditory stimulus were taken from the same set of basic-level categories as reported previously. Additionally, images and sounds of a bear, a chicken, a pig and a seal were included in this second set of stimuli.

The main purpose of Experiment 2 was to identify brain structures related to object processing. In accordance with other studies on visual object processing (e.g. Lerner et al., 2001; Tjan et al., 2006), we intended to compare fMRI signals to intact images of animals on the one hand and scrambled images on the other hand. If no further precautions are taken, scrambling of the color images used in the present experiments might result in a loss or at least in a change of the overall color composition. We used a method developed by Nicolaas Prins (University of Mississippi, USA) which accomplishes scrambling by adding a random phase structure to the existing three phase structures (red, green, blue; RGB) of the original images. Since the relative phases of each layer remained unchanged during phase scrambling, the color composition of intact and scrambled images was essentially the same (see Fig. 1B for example stimuli). Scrambling of the eighteen intact animal vocalizations was performed as described in previous articles from our group (Altmann et al., 2007). Both intact and scrambled sounds had the same average duration of 1040 ms (standard deviation 142 ms, range 734 – 1386 ms), similar to our previous study (Altmann et al., 2007).

All sounds of both experiments were equalized with regard to their root mean square sound intensity. Stimuli were presented with MR-compatible head phones (Resonance Technology, Northridge, USA) using wave-guide tubes for sound transmission. The headphones of the MRI audio system contained fluid- and vinyl-sealed cushions to attenuate ambient noise up to 30 dB. For each subject the applied output volume for the stimulus presentation was adjusted individually to a comfortable level. Stimulus presentation and recordings of behavioral responses was controlled by the software package Presentation (Neurobehavioral Systems, Albany, USA) and triggered by the scanner to ensure synchronous stimulus delivery and fMRI measurements.

## Procedure

For Experiment 1 subjects were administered four runs of the event-related fMRI-adaptation experiment and one run of a high-resolution T1-weighted magnetization-prepared rapid gradient-echo (MP-RAGE) sequence as an anatomical reference. In a separate session, subjects were scanned for at least two runs of Experiment 2, and, again, one run of high-resolution MP-RAGE anatomy for coregistration. This fixed procedure (adaptation first, localizer second) was chosen on purpose because a substantial part

of the stimuli was used in both sessions. Since it is more likely that subtle adaptation effects might be affected by previous contact with the stimuli in a block-design localizer experiment than the other way round, we decided against the counter-balancing of the experimental sequence.

In line with the general rationale of fMRI adaptation paradigms, the experimental conditions of Experiment 1 were created solely by the variation of the second of the two successive AV stimuli (Fig. 1A). For an adapting stimulus S1 and a subsequent stimulus S2 this yielded the following four experimental conditions: 1. the same image and vocalizations for S1 and S2, 2. the same image and a different vocalization, 3. different images and the same vocalization, or 4. different images and vocalizations. Each trial lasted for 10 seconds and started with a short non-stimulus delay of 500 ms and in half of the trials with an additional jitter of one second. The latter was applied to increase the sampling of different subsections of the blood oxygen level-dependent (BOLD) signal and to ensure efficient parameter estimation of the hemodynamic response function (Dale et al., 1999). Next, the first AV stimulus was presented with an average duration of 487.5 ms. After a short interstimulus interval of 500 ms on average, this adapting stimulus S1 was either followed by exactly the same AV stimulus (producing the “same” condition) or a different AV stimulus varying either the auditory, the visual or both components (producing one of the three “different”-conditions). Since stimuli were taken from the same basic-level category such as “dog,” no semantic congruency or incongruency effects were induced either across (from S1 to S2) or within each AV stimulus. Furthermore, due to our randomization procedure each stimulus was equally likely to be presented as S1 and S2.

Each experimental run of the event-related fMRI experiment (Experiment 1) lasted 12 min and 36 s. Four experimental runs per subjects were administered, resulting in 48 trials per experimental condition. Within a single run, 12 trials per condition were presented, intermixed with 12 trials during which only the fixation cross was presented (baseline). Similar to previous studies (Altmann et al., 2004; Doehrmann et al., 2008), the order of presentation was counterbalanced so that trials from each condition were preceded (1 trial back) equally often by trials from each of the other conditions.

During Experiment 2 (localizer) ten successive stimuli were presented in each block followed by a fixation-only interval of equal length. Six different experimental conditions were created involving either intact or scrambled versions of auditory, visual or combined AV stimuli. Each stimulus was presented for one second (in the visual-only conditions) or for the length of the respective sounds (in the auditory-only and the AV conditions; on average 1040 ms). Stimuli were separated by an interstimulus interval of approximately 500 ms, yielding an overall length of each stimulation block and each fixation period of 15 s.

Each experimental condition was repeated two times. Together with an initial fixation each run of Experiment 2 had a length of 9 min and 22 s. The aim was to perform at least two localizer runs per subject. However, all subjects except one were even scanned for three experimental runs.

Throughout both experiments, subjects were instructed to fixate a white cross on a gray background back-projected by an LCD-projector through a mirror fixed to the head coil. In Experiment 1, subjects were furthermore instructed to detect a specific target AV stimulus which was individually assigned to each participant prior to the

first experimental run. Since this target generally referred to one basic-level category, for instance “frog”, subjects could press the button already after the presentation of the first of the two AV stimuli. Thus, the task did not require the detection or identification of one particular exemplar from this particular category. To avoid the contamination with task-related motor activity, the respective target trials were discarded from further analysis of the imaging data. Target stimuli were presented 12 times per experimental run resulting in no substantial loss of experimental trials. This task provided an efficient attentional control compared to passive perception of the stimuli and reduced the impact of task-related activation on the signal patterns measured during the experiment. Subjects’ performance was at ceiling suggesting that stimuli were indeed perceived attentively.

In Experiment 2 subjects had a one-back task requiring a button-press whenever one stimulus was immediately repeated - which was the case in each block of experimental trials. Again, subjects performed at ceiling supporting the view that attentional resources were deployed to stimulus processing. For both experiments, subjects were instructed to use only one hand for responding. The assignment of the left or the right hand was counterbalanced across subjects.

### Imaging and Data Analysis

Imaging for both experiments was conducted using a 3T ALLEGRA (Siemens, Erlangen, Germany) scanner at the Brain Imaging Center, Frankfurt am Main. Similar to recent auditory fMRI studies from our and other labs (Schwarzbauer et al., 2006; Doehrmann et al., 2008) scanning for Experiment 1 was conducted with an interleaved silent steady state EPI sequence to reduce this influence of gradient switching. Each acquisition period lasted 10 seconds and consisted of a silent interval of 6 seconds during which the acoustic stimulus was presented and a data acquisition period of 4 seconds. The following EPI parameters were used:  $TR = 2$  s,  $TE = 30$  ms. For each run 152 volumes consisting of thirty-three axial slices (3 mm thickness with  $3.00 \times 3.00$  mm<sup>2</sup> in-plane resolution, inter-slice gap: 0.6 mm) were collected with a bird-cage head coil. The field of view was 19.2 cm x 19.2 cm with an in-plane resolution of 64x64 pixels. As an anatomical reference we acquired 3D volume scans by using a MP-RAGE sequence with 160 slices ( $TR = 2.3$  s).

For the localizer experiment (Experiment 2) 281 volumes consisting of thirty-four axial slices were collected by using a conventional EPI sequence with the following parameters:  $TR = 2$ ,  $TE = 30$ ; 3 mm slice thickness with  $3.00 \times 3.00$  mm<sup>2</sup> in-plane resolution and an inter-slice gap of 0.3 mm. The field of view was again 19.2 cm x 19.2 cm with an in-plane resolution of 64x64 pixels. Similar continuous sequences were used in our lab both in the auditory as well as in the multisensory domain (Altmann et al., 2007; Hein et al., 2007). The same MP-RAGE sequence as in Experiment 1 was employed for the acquisition of anatomical references.

Functional data were pre-processed and analyzed using the Brainvoyager QX (Brain Innovation, Maastricht, Netherlands) software package. Pre-processing included head movement correction and linear de-trending. Correction of slice scan time using cubic spline interpolation and temporal highpass filtering with 3 cycles per timecourse were only applied to the data of Experiment 2. The 2D functional images were co-registered with the 3D anatomical data, and both 2D and 3D data were spatially normalized into

the Talairach and Tournoux stereotactic coordinate system (Talairach and Tournoux, 1988). Functional 2D data were spatially filtered employing a Gaussian kernel of 8 mm (full width at half maximum; FWHM).

For each experiment, we computed a general linear model (GLM) to estimate the beta weights that model the hemodynamic responses as a combination of two gamma functions (Friston et al., 1998). Whole-brain random-effects (RFX) GLMs across subjects were calculated. For Experiment 1 the model included the following eight predictors: 1) four predictors for the experimental conditions involving pairs of AV stimuli, 2) one predictor for the fixation condition, 3) one predictor for the target condition involving task-related motor activation 4) one predictor for the auditory stimulation resulting from the scanner noise and 5) one predictor for residual T1-related signal decay. For Experiment 2 the GLM was constructed based on the six predictors coding for the experimental conditions.

In a first analysis step we used the GLM from Experiment 2 to define regions of interest (ROIs) with sensitivities to auditory and visual object stimuli, respectively. To this end we computed beta-weight based t-tests contrasting responses to intact versus scrambled stimuli. For the auditory contrast the resulting map was thresholded at  $t(14) = 4.14$ , corresponding to an uncorrected statistical value of  $p < 0.001$ . This map was then projected onto averaged three-dimensional cortex reconstructions generated from the anatomical datasets of all participants (see below) and submitted to cortex-based cluster threshold estimation based on a Monte Carlo simulation with 500 iterations. This procedure resulted in a statistical map with a minimal cluster size of  $94 \text{ mm}^3$  and corresponded to a corrected alpha value of  $p < 0.05$ . Similarly, visual ROIs were generated from a contrast of responses to blocks of intact versus scrambled images. A stricter statistical value was set to  $t(14) = 6.71$  or  $p < 0.00001$  to enhance selectivity of the obtained ROIs. The resulting maps were again projected onto the same averaged cortex reconstruction.

Data from the fMRI adaptation experiment were then analyzed on the basis of these cortex-based auditory and visual ROIs. Effects of auditory and visual repetitions were determined by means of ROI-based RFX GLMs in a manner corresponding to a two-factorial design with the factors “auditory” and “visual”. More precisely, to investigate the impact of auditory changes compared to auditory repetitions we calculated the following contrast, hereafter referred to as the “auditory adaptation contrast”: [(auditory-different, visual-same) + (auditory-different, visual-different) > (auditory same, visual different) + (auditory same, visual same)]. Note that the visual component was completely subtracted yielding an effect which could be attributed solely to responses to auditory stimulation. Analogously, a “visual adaptation contrast” was computed: [(auditory-same, visual-different) + (auditory-different, visual-different) > (auditory different, visual same) + (auditory same, visual same)]. Similar to many fMRI adaptation studies we refer to a significant result in the direction of “different > same” as an effect of “repetition suppression” (RS). In contrast, an effect in the opposite direction, i.e. “same > different”, could be regarded as reflecting “repetition enhancement” (RE; see e.g. Henson et al., 2000). Finally, the interaction term was defined in the following manner: [(auditory-same, visual-same) + (auditory-different, visual-different) > (auditory different, visual same) + (auditory same, visual different)].



Parameter estimates (beta weights) of responses of the auditory and visual ROIs during Experiment 2 were visualized as bar plots in Figure. 2, 3, and 5. These response profiles are reported after the subtraction of the fixation condition. Error bars reflect mean standard errors which were corrected for inter-subject effects similar to previous studies (Altmann et al., 2007; Doehrmann et al., 2008). In particular, the fMRI responses were calculated individually for each subject by subtracting the mean percent signal change for all conditions within this subject from the mean percent signal change for each condition and adding the mean percent signal change for all the conditions across subjects.

We additionally conducted a whole-brain analysis to further corroborate findings from the ROI-based computations. The RFX GLM from Experiment 1 was used to calculate the same adaptation contrasts as previously reported to search for cortical regions exhibiting RS and RE effects. Statistical maps were thresholded to an initial value of  $t(14) = 3.33$ , corresponding to  $p < 0.005$ . Volume-based cluster threshold estimation (Monte Carlo simulation with 1,000 iterations) was applied yielding a cluster size of 5 voxels for the auditory adaptation contrast and 14 voxels for the visual adaptation contrast.

In a final step of analysis we compared on the basis of Experiment 1 responses to changes from the first to the second stimulus (irrespective of modality) to responses to the experimental condition involving no change at all. Similar to the previously described analysis statistical maps were thresholded to an initial value of  $t(14) = 3.33$ , corresponding to  $p < 0.005$ . Volume-based cluster threshold estimation (Monte Carlo simulation with 1,000 iterations) resulted in a cluster size of 6 voxels. No functional clusters survived this correction for the second contrast ("no change" > "change"). Accordingly, results could only be reported on an uncorrected level.

Whenever feasible, group-averaged functional maps were projected onto inflated representations of the left and the right hemisphere of an averaged three-dimensional cortex reconstruction. These reconstructions were generated from the anatomical datasets of all participants by using algorithms from the so-called cortex-based inter-subject alignment procedure (Goebel et al., 2006; Meinenbrock et al., 2007) which were inflated to permit visualization of statistical activation maps both in gyri and sulci. As a morphed surface always possesses a link to the folded reference mesh, functional data can be shown at the correct location of folded as well as inflated representations. Whenever a projection onto these inflated reconstructions was not possible, we used volume-based visualizations on slices of an averaged anatomical dataset from all subjects of our study.

## Results

### **Auditory localizer results and adaptation effects in auditory object-sensitive cortices**

Since we were primarily interested in the response patterns of those cortical regions which are commonly involved in auditory and visual processing of natural objects, we restricted our first analysis to voxels defined by separate localizer runs. A group-based random-effects general linear model (RFX GLM) of Experiment 2 served as the basis

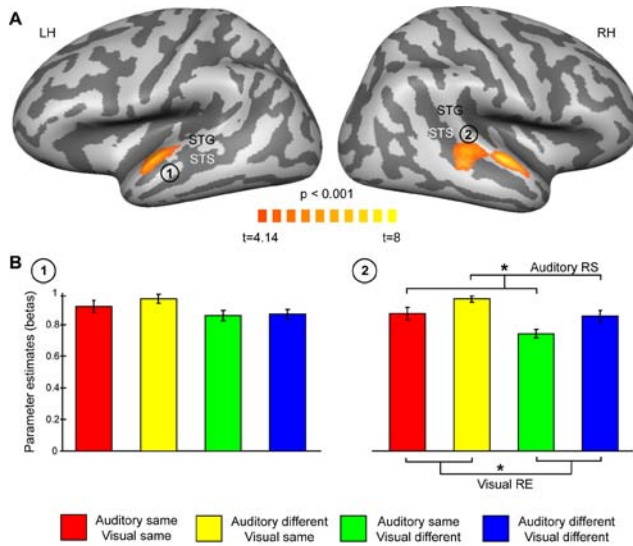


Figure 2. Adaptation effects based on the results of the auditory localizer. (A) Group data ( $n=15$ ) contrasting intact and scrambled auditory stimuli were derived from the general linear model analysis of Experiment 2. Results are shown on lateral views of left and right hemispheres of averaged inflated three-dimensional cortex reconstructions generated from the anatomical datasets of all subjects. Light gray represents gyri, dark gray sulci. (B) Results of Experiment 1 derived from the regions of interests shown in A. Bar plots represent parameter estimates (beta weights) of the four experimental conditions. The responses are reported after subtraction of the beta weights corresponding to the fixation condition on a single-subject basis and error bars represent mean standard errors of the fMRI responses corrected for inter-subject effects. See Methods section for details. Asterisks denote significant ( $p < 0.05$ ) results of contrasts examining repetition effects. The same color conventions as for Figure 1 apply and are additionally provided at the bottom. Abbreviations: LH, RH: left and right hemisphere; RS, RE: repetition suppression and repetition enhancement. STG = Superior temporal gyrus; STS = Superior temporal sulcus.

for the definition of regions of interest (ROIs) in temporal cortex which are particularly related to processing of common auditory objects. Contrasting responses to blocks of intact versus scrambled auditory stimuli (Fig. 2A) resulted in a significant ( $t(14) = 4.14$ ;  $p < 0.001$ ) activation of bilateral superior temporal gyrus (STG) anterior to primary auditory cortex located in Heschl's gyrus (HG) as well as in middle portions of right superior temporal sulcus (STS).

Taking these clusters as ROIs for the investigation of fMRI adaptation effects (Experiment 1) revealed no significant differences as determined by t-tests in ROI-based RFX GLMs between experimental conditions in the left STG (Fig. 2B). However, in the right STG/STS ROI we found significantly stronger responses ( $t(14) = 2.261$ ,  $p < 0.05$ ) to those experimental conditions with changes in the auditory component from S1 and S2 compared to those with auditory repetitions. Significant results for the "auditory adaptation contrast" (see Materials and Methods for details) in the direction "different > same" could be regarded as evidence for auditory repetition suppression (RS).

Significant effects ( $t(14) = 3.467$ ,  $p < 0.05$ ) were additionally found in this region for the comparison of visual repetitions and visual changes. Note that the direction of the



effect was in this case “same > different” which can be considered as visual repetition enhancement (RE). This pattern remained essentially the same when the complete ROI was divided into an anterior and posterior portion. The sole difference was that for the anterior portion visual RE was less pronounced, dropping slightly below significance ( $t(14) = 1.937, p = 0.07$ ). The interaction contrast (see Materials and Methods) was not significant in the auditory ROIs ( $t(14) = 0.897, p > 0.1$  in the left hemisphere, and  $t(14) = 0.863, p > 0.1$  in the right hemisphere).

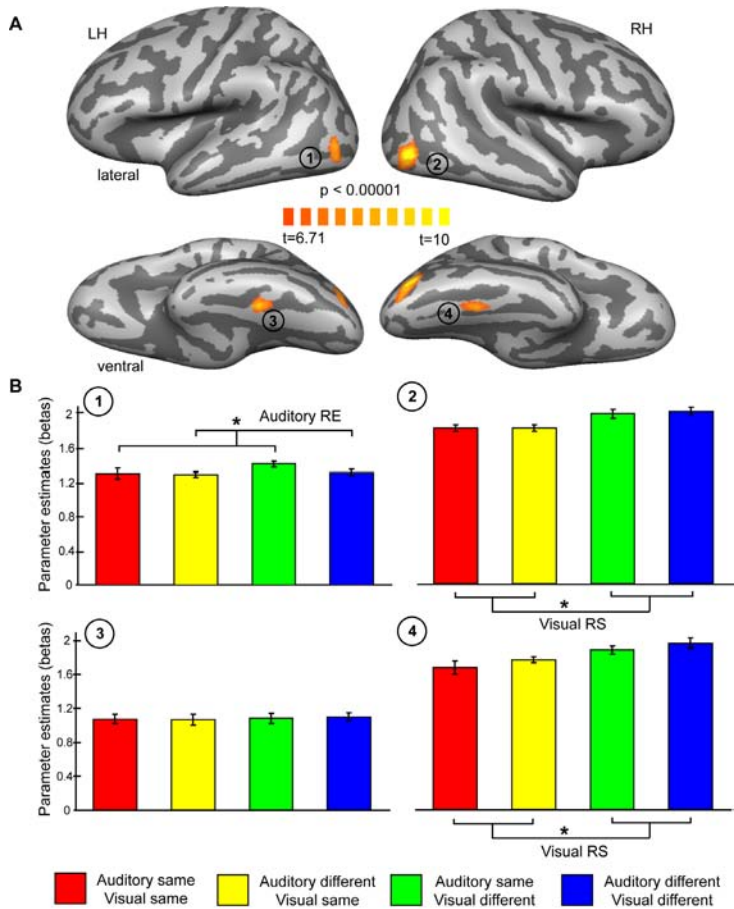


Figure 3. Adaptation effects based on the results of the visual localizer. (A) Group data ( $n=15$ ) contrasting intact and scrambled visual stimuli were derived from the general linear model analysis of Experiment 2. Results are shown on lateral (upper panel) and ventral (lower panel) views of left and right hemispheres of the same averaged cortex reconstruction used in Fig. 2. (B) Effects in the regions of interests shown in A are presented as bar plots of beta weights (again after subtraction of the fixation condition). Error bars represent mean standard errors after correction for inter-subject effects. Color codes are the same as in previous figures and are additionally provided at the bottom. Asterisks denote significant ( $p < 0.05$ ) results of contrasts examining repetition effects. Abbreviations: LH, RH: left and right hemisphere; RS, RE: repetition suppression and repetition enhancement.

### Visual localizer results and adaptation effects in visual object-sensitive cortices

The same group-based RFX GLM of Experiment 2 was applied for the definition of visual ROIs in the lateral occipital (LO) and the ventral temporal cortices. Bilateral regions in LO cortex and fusiform gyrus (FG) were significantly ( $t(14) = 6.71, p < 0.00001$ ) more responsive to blocks comprised of intact than scrambled object images (Fig. 3A). A similar ROI analysis as the one previously described (Fig. 3B) revealed significant RS for the visual adaptation contrast only in right LO ( $t(14) = 2.891, p < 0.05$ ) and right FG ( $t(14) = 3.249, p < 0.05$ ). The interaction was neither significant in right LO ( $t(14) = 0.464, p > 0.1$ ) nor in right FG ( $t(14) = 0.732, p > 0.1$ ). No corresponding effects for visual stimulation were found in left LO and left FG ( $t(14) = 0.814$  and  $t(14) = 0.523$ , respectively). However, in left LO we detected a significantly ( $t(14) = 2.436, p < 0.05$ ) stronger response to auditory repetitions than to auditory changes (RE). Again, no significant interaction as determined by the contrast reported in the previous section was found in left LO ( $t(14) = -0.239, p > 0.1$ ) and left FG ( $t(14) = -0.070, p > 0.1$ ). Details of the ROI analysis in this and the previous section can be found in Table 1.

Table 1 Coordinates and number of significant voxels in regions of interest generated by data from the localizer experiment (Experiment 2)

Region of interest	x	y	z	Number of voxels	Auditory adaptation (Exp. 1)	Visual adaptation (Exp. 2)
<i>Auditory: intact &gt; scrambled</i>						
Left STG	-49	-11	3	747		
Right STG	51	-23	3	1513	RS	RE
<i>Visual: intact &gt; scrambled</i>						
Left LO	-43	-72	0	260	RE	
Right LO	44	-69	-3	581		RS
Left FG	-38	-37	-18	232		
Right FG	36	-41	-18	253		RS

Regions were determined by the group-based RFX GLM of the localizer experiment (Experiment 2) by contrasting intact versus scrambled stimuli for the auditory and the visual sensory modalities. Talairach coordinates are derived from the clusters shown in Figure 2 and 3 and are provided in millimeters. The number of voxels is based on the resolution of the anatomical data set, i.e.  $1 \times 1 \times 1$  mm. Significant effects for the two adaptation contrasts for Experiment 1 were denoted depending on the direction of the respective effect (see text for details). Abbreviations: STG = Superior temporal gyrus, STS = Superior temporal sulcus, LO = Lateral occipital, FG = Fusiform gyrus, RS = Repetition suppression, RE = Repetition enhancement.

### Whole-brain adaptation effects

To further investigate these response patterns beyond the restriction of pre-defined ROIs, we computed the visual and auditory adaptation contrasts also on a whole-brain level (Fig. 4). The visual adaptation contrast revealed a significant ( $t(14) = 3.33, p < 0.05$ ,

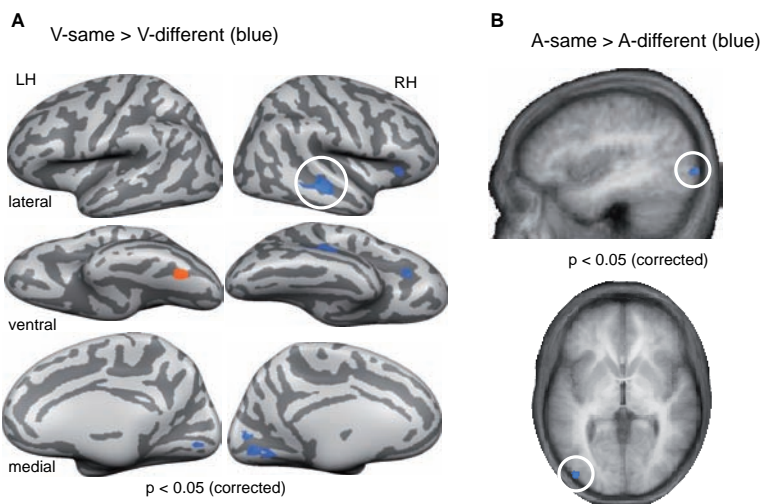


Figure 4. Whole-brain results of the adaptation contrasts. (A) Results for adaptation contrasts examining sensitivities to visual changes (“different” > “same”; map color: orange) and repetitions (“same” > “different”, map color: blue) as indices of repetition suppression and enhancements are shown on lateral (upper panel), ventral (middle panel) and medial (lower panel) views of the same averaged cortex reconstruction used in Figures 2 and 3. (B) A similar contrast as in A was used for the investigation of auditory adaptation effects. Blue color of the map represents stronger effects for auditory repetitions than changes. Data are shown on a sagittal (upper panel) and axial slice (lower panel) of a volume-based anatomical scan averaged across all subjects.

cluster-size corrected) positive effect suggestive of RS in the left FG, posterior to our previously examined ROI (Fig. 4A). An additional cluster in the right FG did not survive cluster-size correction. However, in line with our ROI analysis significant effects in the other direction (“same” > “different”; visual RE) were found in right STG/STS and the right anterior insula. Additional clusters of activation were revealed in bilateral medial occipital cortex, close to early visual areas. Significant positive effects of the auditory adaptation contrast (auditory RS;  $t(14) = 3.33$ ,  $p < 0.05$ , cluster-size corrected) were found only in the precentral gyrus of the left frontal lobe. Again in line with our ROI analysis, significant effects for the reverse contrast (auditory RE) were revealed in a region located in posterior LO of the left hemisphere (Fig. 4B). A summary and details of this analysis are provided in Table 2.

### Whole-brain adaptation effects for global changes and repetitions

We finally asked if there were effects of global changes and repetitions. The former were defined as stronger responses to the mean of those experimental conditions of Experiment 1 which involved changes of any kind (in the auditory or visual or in both modalities) compared to the repetition of stimuli in both components. Computing this contrast of “change > no change” revealed very selective responses ( $t(14) = 3.33$ ,  $p < 0.05$ , cluster-size corrected) in two clusters of the right temporal lobe (Fig. 5): one located in the inferior-temporal gyrus (ITG) and the second more ventral, close to FG. A further ROI-based investigation of these two clusters demonstrated that indeed all experimental conditions involving a change induced significantly stronger responses

Table 2. Whole-brain analysis of Experiment 1

Region of interest	x	y	z	Number of voxels
<i>Visual: different &gt; same (RS)</i>				
Left FG	-37	-54	-15	190
<i>Visual: same &gt; different (RE)</i>				
Right STG/STS	47	-29	0	702
Right anterior insula	32	22	9	161
Right lingual gyrus	3	-69	0	105
Left lingual gyrus	-6	-71	5	125
	-6	-65	0	540
<i>Auditory: different &gt; same (RS)</i>				
Left precentral gyrus	-45	13	9	520
<i>Auditory same &gt; different (RE)</i>				
Left LO	-41	-82	0	220

Regions were determined by the group-based RFX GLM of Experiment 1. Adaptation contrasts (see text for details) for the investigation of repetition effects were calculated on a whole-brain level. Talairach coordinates are derived from the clusters shown in Fig. 4. The same conventions apply as in Table 1.

in the left ITG. In contrast, no significant differences were found in the response of the FG cluster between the experimental condition involving auditory changes only and the one involving no change at all. Thus, this latter region seems to be sensitive only to visual changes.

## Discussion

The present fMRI-adaptation experiment investigated the impact of changes between two successive AV object stimuli without introducing additional variations of congruency. We were able to replicate previous reports of RS for repeated auditory stimuli in the anterior STG/STS and for repeated visual stimuli in LO and FG even under conditions of divided attention. Using fMRI adaptation for the first time in the domain of AV object processing, we additionally demonstrated effects of RE for repeated visual stimuli in the STG/STS and for repeated auditory stimuli in left LO. Thus, cortical regions commonly associated only with processing of unisensory stimuli were found to be modulated in their activity for repeated stimuli in the respective other sensory

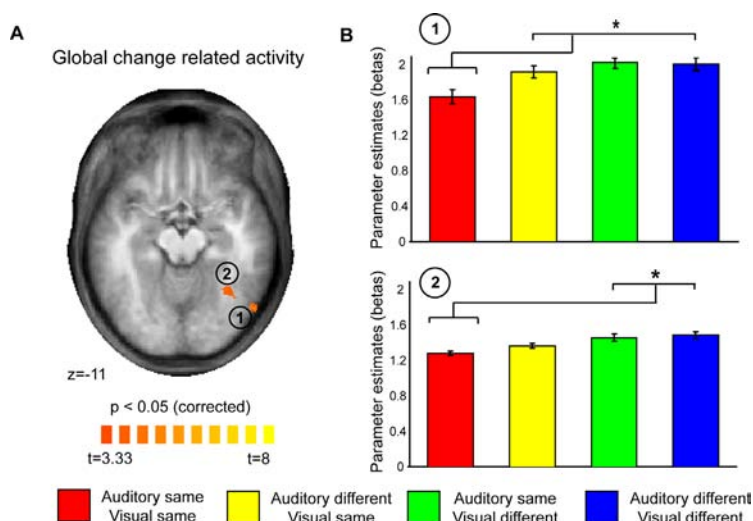


Figure 5. Whole-brain global change- and repetition-related effects. (A) Global change-related effects were determined on the basis of results from Experiment 1. Mean responses to all three experimental conditions involving a change in the auditory, visual or both components were compared to the responses of a repeated stimulus. Significant effects are shown on an axial slice of the same averaged anatomical dataset used in Figure 4. Cluster #1 is located on the inferior temporal gyrus ( $x=51$ ,  $y=-61$ ,  $z=-7$ , size: 366 voxels) and cluster #2 on the fusiform gyrus ( $x=34$ ,  $y=-54$ ,  $z=-11$ , size: 222 voxels). (B) Barplots of beta weights derived from Experiment 1 are shown for the cortical regions marked in A. Color codes are provided at the bottom. Each experimental condition involving a change in one or both sensory modalities was separately compared by means of t-test to the experimental condition involving no change for the regions determined in A. Asterisks denote significant effects ( $p < 0.05$ ) of t-tests on beta weights of Experiment 1. Error bars of plots were corrected for inter-subject effects.

modality. We start with a brief examination of our RS effects and then proceed with a discussion of the novel findings of RE in AV object processing. The last part will be on effects found in the present fMRI-adaptation study in ITG which showed a release from adaptation for any experimental condition involving a change compared to the “no change” condition.

### Effects of auditory and visual repetition suppression

The most conspicuous finding from our study is a differential pattern of RS and RE effects. The former are in good correspondence with previous results obtained by fMRI-adaptation or priming paradigms. In two earlier studies we were able to demonstrate suppressive effects for animal vocalizations in anterior portions of the STG (Altmann et al., 2007; Doehrmann et al., 2008). These could not be explained by non-selective adaptation to stimuli of a certain frequency complexity (Altmann et al., 2007). Furthermore, when compared to stimuli of a category of man-made objects, adaptation in STG did not occur for these stimuli but appeared to be selective for animal vocalizations (Doehrmann et al., 2008). These findings were also in line with results from studies using fMRI (Bergerbest et al., 2004) and ERPs (Murray et al., 2006; 2008). Together, these findings

underscore that the present experiment indeed tapped into processes related to the activation of a ventral auditory “what”-stream.

Similarly, visual RS effects for common objects have been previously demonstrated in right LO and FG by a large set of studies using fMRI adaptation techniques. More precisely, LO is generally sensitive to visual object stimuli compared to patterns of elements of similar complexities which do not form a complete object (Malach et al., 1995). RS was measured both for the repetition of real and for nonsense objects (Vuilleumier et al., 2002) further supporting a relative invariance of this cortical region to the type of visual stimuli in use. Other studies revealed that activity in this cortical region was more related to the actual percept of the subject and less to the processing of physical features of an object (Kourtzi and Kanwisher, 2001; Weigelt et al., 2007b). In contrast, FG might operate on a higher processing level, likely to be connected to semantic aspects of visual stimuli. This view is supported by a general preference for real compared to nonsense objects (Vuilleumier et al., 2002) and different cortical representations for various stimulus categories both within and in close proximity to FG (Downing et al., 2006).

From a more general perspective we conclude that our stimuli indeed engaged the recognition-related streams. The effects of RS in cortical regions related to auditory and visual object processing are no trivial result since subjects were instructed to attend both sensory modalities. Several studies demonstrated a close interrelation of attention and fMRI adaptation with a decrease or even lack of RS when attention was not directed to the repeated stimuli (Murray and Wojciulik, 2004; Weigelt et al., 2007a). Therefore, since we were able to measure RS effects, the division of attention did not undermine typical object processing as examined in classical unimodal fMRI adaptation experiments.

### **Auditory and visual repetition enhancement**

The previously discussed effects are consistent with the view that repetitions within sensory modalities indeed induced typical adaptation effects. However, more important for research on sensory integration is the question of repetition effects across sensory modalities. Interestingly, in the present study these were found as enhancement rather than suppression effects (see the following section for a discussion of cross-modal suppression effects).

In contrast to the widely-reported effects of RS, RE is a less-discussed phenomenon although it was present even in studies utilizing single-cell recordings in non-human primates (see e.g. Li et al., 1993). In human subjects, effects of RE were found both in fMRI and electrophysiological studies focusing on oscillatory activity. For instance, Henson and colleagues (2000) examined the influence of stimulus familiarity on repetition priming while measuring fMRI. Faces (famous or non-famous) and symbols (familiar or abstract) both induced effects of RE in FG of the right hemisphere. Similarly, induced oscillatory activity in the gamma band showed effects of RE for unfamiliar stimuli such as pseudowords (Fiebach et al., 2005) or nonsense objects (Gruber and Müller, 2005). Finally, these effects are not restricted to the visual sensory modality, but have been described also in the auditory domain (Gagnepain et al., 2008).

Given these previous reports it is not surprising that the present study also revealed effects of RE. However, the important question is what these enhancement effects re-

flect on a functional level. For unfamiliar material it has been suggested (Henson, 2003; Gruber and Müller, 2005) that RE is a signature of memory encoding and therefore the generation of a novel cortical representation for these stimuli. A stimulus repetition might enhance activity in a population of neurons coding for the previously presented stimulus. This neuronal enhancement is then translated into a BOLD effect as measured by fMRI. Another, not necessarily contradicting explanation of RE is based on studies which systematically varied the degree of stimulus salience (Kourtzi et al., 2005; Turk-Browne et al., 2007). Presentation of stimuli which were previously learned either with or without a noisy background resulted in stronger fMRI responses for the former in several regions of the ventral visual stream upon subsequent presentations (Kourtzi et al., 2005). A similar effect could be obtained for the place-sensitive cortical region in ventral occipito-temporal cortex for the immediate repetition of two low-contrast scenes (Turk-Browne et al., 2007). In the latter case the second stimulus “aids” the processing of the suboptimal first stimulus thereby inducing repetition enhancement.

Our RE effects in right STG/STS and in left LO can similarly be interpreted along these lines. An auditory region like right STG/STS reacts with a stronger signal for auditory changes (auditory release from adaptation, RfA) than for auditory repetitions. However, a visual stimulus accompanying the auditory stimulus could be regarded similar to the low-contrast stimuli in the study by Turk-Browne et al. (2007) also as suboptimal for this cortical region. If neurons in this region were completely insensitive to the content of the visual stimulus, we would have found no difference for repeated versus different visual stimuli. Most interestingly, this was not the case. In contrast to this, we detected a stronger activation of right STG/STS for visual repetitions compared to visual changes possibly also due to an increased processing depth for a second presentation of the suboptimal visual stimulus. Conversely, in left LO the strongest responses were measured for visual changes (visual RfA) and auditory repetitions (green bars in Fig. 3). In the case of LO, auditory stimuli could be regarded as suboptimal and therefore produced enhanced responses after stimulus repetitions.

Drawing together these ideas, our major findings can be described as follows (Fig. 6). The most important “driving forces” for the cortical responses are mechanisms of RS and RfA. The most pronounced responses in right STG/STS were induced by auditory changes and a corresponding strong auditory RfA which is likely to be further modulated by RE due to visual stimulus repetitions. At the same time, these repeated visual stimuli suppressed the response in extrastriate regions such as LO. Furthermore, although the underlying mechanisms are less clear, suppressive effects might also be induced in these regions by auditory changes (Fig. 6A). For left LO the major “driving force” is a visual RfA by visual changes. This is further increased by an accompanying auditory RE due to the repetition of sounds. However, these repeated stimuli produce at the same time an auditory RS in cortical regions such as STG/STS (Fig. 6B). Note that the model presented in Fig. 6 is oversimplified to some extent since only in STG/STS all mechanisms of RS, RfA and RE operate in conjunction while these different processes seem to be more distributed in visual cortex with more pronounced effects of RS and RfA in more ventral regions such as FG and effects of auditory RE in left LO.

These putative interactions between higher-order sensory cortices are in line with a growing body of recent literature supporting the view that even cortical regions previously regarded as strictly unisensory could at least be modulated by stimuli



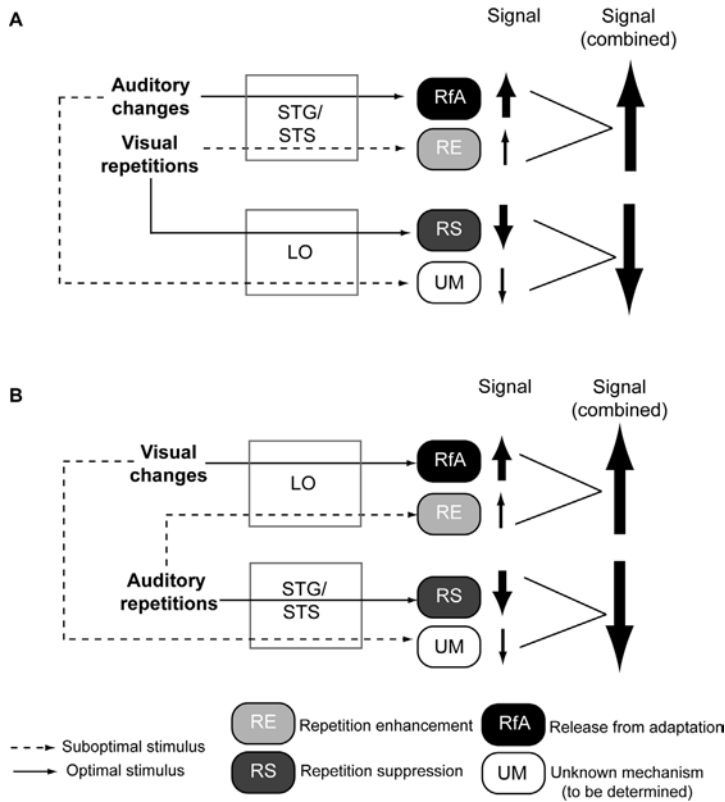


Figure 6. Schematic illustration summarizing the major results of the present study. Functional profiles of higher-order auditory (STG/STS) and visual (LO) regions are illustrated under the conditions involving (A) auditory changes and visual repetitions and (B) visual changes and auditory repetitions. A legend of abbreviations used is provided at the bottom of the figure. Upward and downward arrows symbolize signal increases and decreases, respectively, while the thickness of the arrows illustrates the strength of these changes.

presented in the respective other sensory modality. Recordings in auditory cortices of non-human primates have revealed both enhanced and suppressed activity for AV stimuli such as monkey faces and vocalizations (Ghazanfar et al., 2005) which are likely to result from interactions between these regions and multisensory cortices such as STS (Ghazanfar et al., 2008). Also, tracer studies in these primates (Falchier et al., 2002; Clavagnier et al., 2004) suggest direct connections between auditory and visual cortices without additional connections onto mediating neurons in multisensory regions. Additional work investigating projections in the cat's sensory cortices suggests even direct connectivity between auditory and higher-order visual regions (Clemon et al., 2008). In humans, neuroimaging studies revealed influences of visual stimuli on auditory cortices (e.g. van Atteveldt et al., 2004; Hein et al., 2007) and of auditory stimuli on visual cortices (e.g. Giard and Peronnet, 1999; Amedi et al., 2007).

However, at least to our knowledge no study to date has found similar repetition-induced modulatory influences of concurrently presented AV stimuli on the processing of sensory cortices without introducing additional variations with respect to seman-



tic congruency or by comparing responses to stimuli with different sensory complexity (e.g. bimodal versus unimodal stimulation). In contrast, even when controlling for these different factors our study demonstrates that repeated stimuli in one sensory modality can enhance the activity in the respective other sensory cortices.

### **Global repetition suppression effects and alternative accounts**

The final set of findings of the present study concerns what we have termed “global” RS effects. We found such effects of enhanced fMRI responses in the right FG and more laterally in right ITG for experimental conditions involving any change (auditory, visual, or both) compared to the condition with no change. Interestingly, these two regions differed with regard to their sensitivity to particular changes. The FG showed only stronger responses for experimental conditions with a visual change and was therefore primarily driven by visual RS and RfA. This pattern was already discussed in the section on visual RS. In contrast, the ITG additionally exhibited more pronounced activity even when a change occurred in the auditory component. We will therefore focus now on the latter cortical region.

Several previous crossmodal neuroimaging studies reported activity in the ITG, particularly in response to visual and auditory stimuli. For instance, Tranel and coworkers (2005) found ITG to be involved during naming tasks on stimuli presented either in the visual or the auditory sensory modality. The authors suggested that this cortical structure mediates conceptual knowledge retrieval and the formation of a word form during naming. However, the effects found by Tranel et al. (2005) were confined to the left hemisphere possibly due to the linguistic task, whereas the present study revealed only responses in right ITG. Similar to these reports of Tranel et al. (2005), some repetition priming studies (e.g. Buckner et al., 2000; Cohen et al., 2004) found signal reductions in left ITG for repeated word generations cued either in the auditory or in the visual modality. As in the case of Tranel and coworkers (2005) this finding has been discussed particularly in relation to semantic aspects which were rather detached from perceptual processing, i.e. operating on an “supramodal” level irrespective of which sensory modality was used for introducing information into the system.

While these findings are interesting in their own right, they can also provide an answer to a possible objection to our interpretation of our major findings as presented in the previous section. More precisely, an important issue in this respect could be the influence of attention. For instance, Shomstein and Yantis (2004) showed that shifts of attention during AV stimulus presentations between auditory and visual streams result in a particular pattern of suppression and enhancement of activity in the respective sensory cortices. However, the response pattern measured in the present study could not easily be accounted for by these attentional effects. Shomstein and Yantis (2004) found more pronounced fMRI responses in fronto-parietal regions for those experimental conditions involving changes (or “shifts of attention” in their terminology) than for those with no changes. However, in the present study, this corresponds exactly to the contrast comparing responses to any change to the condition involving no change at all (Fig. 5A). When computing this contrast we did not find fronto-parietal regions, rendering it at least unlikely that covert shifts of attention could completely account for our pattern of results.

Furthermore, this attentional account can offer no straightforward answer to the response pattern revealed in the present study. One would have to argue that changes in one sensory modality (from the first to the second AV stimulus) caused a shift of attention to that modality and a concomitant enhancement of activity in the respective sensory cortex and suppressive effects in the other sensory cortex. As explained before, this approach could not fully explain why particular experimental conditions involving changes in the respective other sensory modality caused a stronger enhancement than other conditions. In contrast, we think that our approach as described in the previous section can provide a more extensive account for the response patterns found in the present study.

## **Conclusion**

The present study used an fMRI-adaptation paradigm to investigate how cortical structures were modulated by changes in the auditory, visual, or both components of two successive AV stimuli. Even when controlling for stimulus complexity and semantic congruency our design could reveal subtle influences of (repeated) visual stimuli on processing in higher-order auditory cortices and of (repeated) auditory stimuli on extra-striate visual cortex. These effects could be explained by the complex interaction of different repetition-related mechanisms which can only be revealed in an fMRI-adaptation paradigm such as the one employed here. Thus, our novel findings are not only interesting in their own right, but also support the conclusion that fMRI adaptation can indeed be fruitfully applied to the domain of AV integration. We therefore expect that this method will soon be utilized in a widespread manner to further unravel the integrative mechanisms working in the human brain.

## **Acknowledgements**

This study was supported by the Bundesministerium für Bildung und Forschung (Brain Imaging Center Frankfurt, DLR 01GO0508), the Frankfurt Medical Faculty intramural young investigator programme (to M.J.N.) and the Deutsche Forschungsgemeinschaft (DFG AL 1074/2-1). The authors would like to thank Tim Rohe for assistance during cortex reconstructions and Benjamin Rahm for helpful discussions.



# 6

## **Semantics and the multi-sensory brain: How meaning modulates processes of audio-visual integration**

By using meaningful stimuli, multisensory research has recently started to investigate the impact of stimulus content on crossmodal integration. Variations in this respect have often been termed as “semantic”.

In this paper we will review work related to the question for which tasks the influence of semantic factors has been found and which cortical networks are most likely to mediate these effects. More specifically, the focus of this paper will be on processing of object stimuli presented in the auditory and visual sensory modalities. Furthermore, we will investigate which cortical regions are particularly responsive to experimental variations of content by comparing semantically matching (“congruent”) and mismatching (“incongruent”) experimental conditions. In this context, recent neuroimaging studies point toward a possible functional differentiation of temporal and frontal cortical regions, with the former being more responsive to semantically congruent and the latter to semantically incongruent audio-visual (AV) stimulation.

To account for these differential effects, we will suggest in the final section of this paper a possible synthesis of these data on semantic modulation of AV integration with findings from neuroimaging studies and theoretical accounts of semantic memory.

Doehrmann, O., and Naumer, M.J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research*, 1242, 136-150.

## Introduction

It is common knowledge that we experience our environment by means of different sensory channels. However, while it seems natural for the perceiver that information from several sensory modalities is integrated to form a coherent whole, a central question in the behavioral and neurosciences concerns the mechanisms mediating this integration. Several factors have been reported to have a major impact on these integrative mechanisms such as spatial and temporal correspondence of incoming signals, e.g. from the visual and auditory channels (Stein and Meredith, 1993). While the influence of these factors can be explained by the receptive-field properties of multisensory neurons in subcortical and cortical structures, other mechanisms of integration like a close correspondence of content operate on a higher level and are more difficult to accommodate in theoretical frameworks.

The latter mechanisms have frequently been called “semantic” (e.g. Calvert, 2001; Laurienti et al., 2003). Typical perceptual situations involving such a semantic context are those in which a complex visual stimulus is paired with a matching auditory counterpart, for instance the picture of a dog and a corresponding barking sound. Both psychophysical (e.g. Laurienti et al., 2004) and neuroimaging studies (e.g. van Atteveldt et al., 2004; Hein et al., 2007) have intended to capture these situations by contrasting “congruent” and “incongruent” experimental conditions. It has become evident that even when only the crossmodal domain is considered, congruency could be modulated in a number of ways. For instance matching and mismatching conditions could be generated for various types of content such as emotional expressions in face and voice stimuli (Dolan et al., 2001), variations in speaker identity (Schweinberger et al., 2007) or images and sounds of common objects (Noppeney et al., 2007).

Given this wide range of different contexts it is not easy to provide an exact definition of “semantics” as it is used in the research field of multisensory integration. To a first approximation, a very weak degree of “semantic” variation might already be given whenever stimulus properties other than spatial and temporal factors, for instance shape or pitch, are varied in experimental contexts. However, most studies relevant here used more complex visual and auditory object stimuli, for which associations had been established in the course of an individual’s learning history. In many cases, these associations can be regarded as “arbitrary” by connecting information conveyed in different sensory modalities, in contrast to features such as motion or rhythm which could in principle be perceived in more than one sensory modality (sometimes called as “amodal” stimulus properties in developmental contexts, see e.g. Lickliter and Bahrick, 2004). However, it is likely that while arbitrarily related stimuli from different sensory modalities become associated, they increasingly contribute to a concept referring to a certain object. Thus, semantics in a stronger sense possibly involves a whole network of information related to objects and their diverse properties and is therefore heavily connected to semantic memory (Martin, 2007; Patterson et al., 2007).

In this review, we will focus on this stronger sense of semantics and its impact on the integration of stimuli presented in different sensory modalities. Moreover, we will cover exclusively work related to audio-visual (AV) integration since these two sensory modalities most reliably transmit meaningful content and, conversely, their integration is more likely to be influenced by semantics (however, see Gottfried and Dolan (2003)

for a less typical example from the visual-olfactory domain). In a first step we will discuss psychophysical studies investigating the impact of semantics in tasks involving AV stimulation to underscore the view that semantic associations are indeed causally relevant for a subject's behavior, especially in multisensory contexts. The major part of this review will then be devoted to the question of which cortical structures actually do exhibit AV integration effects induced by semantic modulations. While in most cases, effects for AV integration have been investigated by comparing the responses to bimodal stimulation with those to unimodal stimuli, we will comment on the continuing debate on appropriate criteria for integrative effects in neuroimaging studies in section 3 ("Cortical networks for AV integration of common objects").

For the factor "semantics," also so-called "congruency effects" obtained by the comparison of conditions involving semantically matching ("congruent") and non-matching ("incongruent") AV bimodal stimuli will be of particular importance. Findings for common AV objects obtained by functional magnetic resonance imaging (fMRI) and noninvasive electrophysiological methods such as electroencephalography (EEG) and magnetoencephalography (MEG) will then be further compared to results from recent studies using other types of stimuli such as AV speech to determine the degree of specificity of the cortical networks. In the final section we will suggest a possible synthesis of how cortical regions involved in AV integration may possibly interact with regions associated with semantic memory processes.

## **Behavioral effects of varying semantics during AV stimulation**

A large number of behavioral studies has revealed impressive facilitative effects of concurrent crossmodal stimulation, for instance during simple reaction time tasks (Forster et al., 2002) and more complex target detection (Molholm et al., 2004), with respect to detection thresholds (Frassinetti et al., 2002) or visual perceptual learning (Seitz et al., 2006). Thus, it is widely accepted that stimuli presented in different sensory modalities could provide complementary information yielding elevated performance levels likely to be caused by neural integration processes instead of simple probability summation (Miller, 1982).

While a large number of studies has investigated the influence of temporal and spatial factors on multisensory integration (e.g. Sugita and Suzuki, 2003; Fujisaki et al., 2004; Navarra et al., 2007), the impact of semantics on these integration processes has been investigated by comparatively few studies. Laurienti and colleagues (2004) developed a behavioral paradigm which allowed them to examine effects of congruent and incongruent AV stimuli compared to a control condition involving congruency in a visual within-modality context. Significantly faster response times were found selectively for congruent AV pairings compared to unimodal stimulation, but not for semantically congruent, visual intra-modal conditions. In contrast, reaction times were significantly longer for incongruent conditions in both crossmodal and intra-modal contexts. This study therefore already highlights the importance of employing appropriate intra-modal control conditions to demonstrate specifically multisensory effects of semantic congruency. However, effects with respect to incongruency seem to be less specifically bound to crossmodal stimuli.

However, behavioral facilitation effects for semantically associated stimulus components might also be modulated by directing attention to a single sensory channel. Recently, interaction effects of semantic congruency and sensory modality were reported for a name-verification task on natural objects, demonstrating stronger effects of congruency when the auditory modality was attended (Yuval-Greenberg and Deouell, 2007). These effects suggest a more pronounced influence of concurrent visual stimulation on auditory perception than vice versa, possibly due to the fact that in most cases the visual modality provides more reliable information for object-recognition (see also the so-called “modality-appropriateness hypothesis” proposed by Welch and Warren, 1986).

Recent studies have started to investigate the impact of semantic congruency on previously described multisensory perceptual effects. For instance, Koppen et al. (2008) examined if stimulus congruency significantly influences an effect during a speeded response task in which participants had to determine which sensory modality had been stimulated (auditory, visual or AV). Typically, subjects fail more often to respond to the auditory component of an AV stimulus than to the visual one (Colavita, 1974), again suggesting visual dominance. Interestingly, the magnitude of the so-called Colavita effect was unaffected by semantic associations between AV animal stimuli. Moreover, during congruent AV stimulation subjects were significantly less accurate and slower than during incongruent conditions. However, it should be noted that the effects found by Koppen et al. (2008) during speeded modality-determination tasks might be less likely to benefit from semantic congruency than tasks requiring a target detection (Laurienti et al., 2004; Molholm et al., 2004). For instance, the identification of a pre-defined target stimulus often requires some advanced level of object processing which could be enhanced by redundant information from a second modality. In contrast, the determination of the stimulated modality could, in principle, also be accomplished for stimuli lacking any semantic content or, as demonstrated by Koppen et al. (2008) might even be negatively affected by concurrent, semantically matching stimuli in a second sensory modality.

Effects of semantic congruency have also been reported during AV temporal order judgment tasks, i.e. when subjects had to determine which modality had been stimulated first. Using AV speech stimuli, Vatakis and Spence (2007) demonstrated that subjects found such a task easier when both unimodal components matched (in this case: gender of voice and face). Similarly, Gallace and Spence (2006) were able to show that a match even of very simple stimuli (disks of varying sizes and simple tones with high or low frequency) can have a significant influence on a size-judgment task for delayed presentations of disk stimuli – even if subjects were explicitly instructed to ignore the auditory stimuli. Thus, congruency effects might already be present for low-level stimuli. Furthermore, processing of high-level visual stimuli such as faces could also be significantly influenced by low-level stimuli such as pure tones. For instance, Smith et al. (2007) recently reported that a gender-classification task on androgynous face stimuli could be significantly biased depending on the concurrent presentation of pure tones either in the male or female fundamental-speaking-frequency range. Importantly, these tones transmitted no complex spectral information generally associated with human voices and, therefore, are typically processed on a very early level of auditory cortex. Thus, under certain circumstances the combined presentation of AV

stimuli which are congruent even on very subtle levels could significantly determine perceptual decision processes and thereby behavioral performance.

Finally, concurrent presentation of semantically matching AV stimuli does not only enhance immediate performance, but also has some facilitative effects on later memory performance. Murray et al. (2004) showed that recognition performance (old/new decisions) for repeated presentations of line drawings displaying common objects was significantly improved if these pictures were paired with task-irrelevant, but semantically matching sounds during the initial presentation. In a follow-up study, Lehmann and Murray (2005) demonstrated that these facilitative effects were not simply caused by the presence of any multisensory encoding context, since recognition accuracy was unaffected for object images which had been paired with a simple tactile stimulus during initial presentations. Interestingly, pairing a novel object image with an auditory stimulus lacking semantic content, i.e. a pure tone of 1000 Hz, induced a reduction of accuracy on subsequent presentations of these visual stimuli. More importantly, only a semantically congruent, but not incongruent initial AV presentation had a positive effect on subsequent recognition performance during repeated image presentations. Thus, memory representations of previously encountered visual objects seem to specifically benefit from semantically matching auditory stimuli possibly due to a more elaborated encoding leading to a richer or more robust stimulus representation. However, it should be noted that the degree of specificity of these effects to the multisensory domain could not be fully explored since the authors did not employ intra-modal control conditions (Laurienti et al., 2004).

In summary, a number of recent studies was able to demonstrate significant influences of semantically related AV stimuli on behavioral effects. In many cases semantically congruent compared to incongruent AV stimulation substantially improved the performance. In the following sections we turn to the question which cortical networks form the most likely neural basis for these integrative effects.

## **Cortical networks for AV integration of common objects**

As reviewed above, psychophysical studies using semantically modulated AV stimulation point toward behavioral advantages for congruent compared to incongruent conditions. In this section we will focus on recent neuroimaging studies using AV stimuli of common objects to investigate the interaction of integrative and content processing. While our overall aim is to identify candidate regions in human cerebral cortex which are particularly sensitive to semantic variations in AV objects, we will first focus on fMRI work and thereafter on electrophysiological findings.

### **fMRI studies**

As already mentioned in the introduction, effects concerning multisensory integration have mostly been investigated by comparing responses to bimodal stimuli with those to unimodal stimulation. Some commonly reported cortical activations for AV stimulation compared to unimodal control stimuli are depicted schematically in Figure 1 (for details see e.g. Ghazanfar and Schroeder, 2006; Kayser and Logothetis, 2007; Driver and Noesselt, 2008).



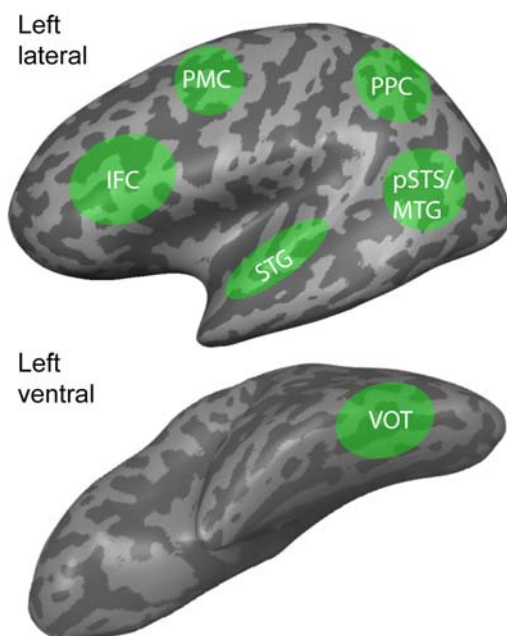


Figure 1. Schematic illustration of commonly reported regions involved in AV integration. An inflated three-dimensional cortex reconstruction of the left cerebral hemisphere (MNI template brain) is shown in lateral (upper panel) and ventral (lower panel) views. Gyri and sulci are shown in lighter and darker shades of grey, respectively. The selected example studies reported AV effects in the following regions: inferior frontal cortex (IFC; Hein et al., 2007) premotor cortex (PMC; Skipper et al., 2007), posterior parietal cortex (PPC; Noppeney et al., 2007), superior temporal gyrus (STG; van Atteveldt et al., 2004), posterior superior temporal sulcus/middle temporal gyrus (pSTS/MTG; Beauchamp et al., 2004; Hein et al., 2007), ventral occipito-temporal cortex (VOT; Beauchamp et al., 2004).

Regarding fMRI data the question of appropriate statistical criteria to determine multisensory integration effects has been continuously discussed (see e.g. Calvert, 2001; Laurienti et al., 2005; Beauchamp, 2005). Based on invasive electrophysiological findings it has been argued that also in fMRI studies multisensory integration might be indicated by the response to AV stimuli exceeding the linear sum of the responses to unimodal stimulation, i.e.  $AV > A + V$  ("superadditivity"; see e.g. Calvert et al., 2001). However, in many contexts this criterion might be too strict, especially if the stimuli are not presented in a degraded manner or close to threshold (Stanford and Stein, 2007). Accordingly, we will also consider fMRI studies employing criteria based on "subadditivity" (sometimes also called the "max"-criterion; Beauchamp, 2005) in which responses to AV stimuli have to exceed the most pronounced unimodal response, i.e.  $AV > \max(A, V)$  (see e.g. Hein et al., 2007). Finally, with respect to the question of sensitivity to semantic modulations in a given cortical region, we will particularly focus on direct bimodal comparisons of semantically congruent and incongruent AV stimuli. A possible advantage of this approach is a similar degree of perceptual complexity (due to bimodal stimulation in both conditions) together with a variation in semantic re-

latedness of the stimuli. Thus, studies employing unimodal and bimodal experimental conditions as well as AV congruency modulations are particularly relevant for the main purpose of this review.

One of the first fMRI studies to investigate the impact of semantics on AV integration (Laurienti et al., 2003) presented semantically congruent and incongruent AV stimuli of common objects while subjects performed a task only on the visual stimuli. The authors found regions in anterior cingulate and medial prefrontal cortices activated more strongly during congruent compared to incongruent AV stimulation. However, the interpretation of these results was difficult for several reasons. First, attention was directed to the visual stimuli while auditory stimuli had to be ignored. Other studies (Alsius et al., 2005; Degerman et al., 2007) suggested that both components of an AV object had to be attended for crossmodal integration to occur. Thus, task-related and perceptual effects were difficult to disentangle in that study. Second, the authors neither employed a baseline condition nor any unimodal control conditions. While the former might have allowed ruling out effects generated by the comparison of BOLD responses below baseline for both conditions, the latter could have enabled assessing if integrative effects were actually present in these regions (by contrasting AV conditions with the respective unimodal controls). Finally, in order to increase the overall statistical power Laurienti et al. (2003) did not differentiate between object categories. However, a possible finding of category-related AV integration effects would have represented strong evidence for the impact of semantic content in a crossmodal experimental setting.

Later fMRI studies intended to overcome these limitations. Beauchamp et al. (2004) conducted a series of experiments using AV stimuli taken from animate (animals) and inanimate (tools) categories. Particularly cortical regions comprising the posterior superior temporal sulcus (pSTS) and posterior middle temporal gyrus (pMTG) as well as the ventral temporal cortex suggested effects of AV integration as evidenced by stronger fMRI signals to combined AV stimulation than to each of the unimodal stimuli alone. Furthermore, these effects for pSTS/MTG seemed to reflect the early perceptual phase involving AV object stimuli rather than the behavioral phase related to an identification task. This was in contrast to regions of the dorsolateral prefrontal cortex (DLPFC) which responded more strongly during the task phase. Interestingly with respect to semantics, Beauchamp and colleagues found only a trend towards statistical significance for enhanced responses to congruent (compared to incongruent) AV stimuli, although this lateral temporal region seemed to be more sensitive to intact than scrambled stimuli (presented in the visual and auditory modalities, respectively). Taken together, this seminal fMRI study revealed a fronto-temporal network of regions that is likely to mediate integration of complex, meaningful AV object stimuli. This interpretation has been further corroborated by several recent fMRI studies. For pSTS/MTG, robust effects were induced by object-related AV stimulation, which could even exceed the linear sum of the fMRI responses to isolated unimodal stimuli (a “superadditive” effect; see Stevenson et al., 2007). This further supports the assumed role of this region in AV object recognition (see also Sestieri et al., 2006).

However, both the exact functional role of the regions constituting the described fronto-temporal network and the impact of semantics on AV integration as suggested by significant effects of semantic congruency have still remained to be determined.

A first step to address these open questions might be the differentiation of cortical regions preferring congruent or incongruent AV object stimuli (Fig. 2). In line with this reasoning, Belardinelli et al. (2004) demonstrated effects of semantically congruent (compared to incongruent) AV object stimulation in the left ventral and medial temporal cortex as well as in the lingual gyrus of both cortical hemispheres, while in particular regions of the left inferior frontal cortex (IFC) were more active for incongruent conditions. In contrast, Taylor et al. (2006) reported a trend towards more pronounced fMRI responses for incongruent (compared to congruent) AV stimuli as well as to animate (compared to inanimate) stimuli in the perirhinal cortex of the medial temporal lobe. However, none of these effects were present in the pSTS/MTG possibly suggesting a rather “presemantic” role of this cortical region. Consistent with other reports, Taylor et al. (2006) additionally found stronger activations during incongruent AV conditions (compared to unimodal controls) in bilateral medial frontal cortex (MFC) as well as the anterior cingulate cortex (ACC). Both cortical structures have been repeatedly implicated in the detection of conflicts between simultaneously active, competing representations of stimuli both within and across sensory modalities and error-related processing (see e.g. Botvinick et al., 2004; Weissman et al., 2004). Conflict detection in the ACC possibly engages the DLPFC resulting in increased cognitive control in service of conflict resolution (Carter and van Veen, 2007). Accordingly, activation of these regions does not necessarily reflect processes related to crossmodal conflict, underscoring the need for appropriate control conditions, for instance, by the induction of incongruency effects also in the unimodal domain (Taylor et al., 2006). However, others (Naghavi et al., 2007) failed to find effects for incongruent AV stimulation (reporting only effects for congruent AV object stimuli in the right claustrum/insula) which highlights the requirement of further studies in this domain of multisensory research.

In one approach to address this question, Noppeney and colleagues (2007) have recently employed a priming paradigm to further elucidate the sensitivity of cortical regions to incongruent auditory and visual stimulation. The authors used either pictures (of animals or man-made objects) or words as visual primes and studied their respective impact on the processing of subsequently presented natural sounds or spoken words. Generally, the presentation of semantically incongruent (compared to congruent) prime-target combinations produced activation in a network of regions in the temporal, parietal and frontal lobes. However, these regions did not participate to the same degree in the processing of different types of incongruent stimuli as employed in this study. Only the medial and inferior frontal cortex showed incongruency effects both for sounds and for spoken words consistent with its previously discussed role in conflict monitoring and error processing. Relatively stronger incongruency effects for words were found in the left pSTS/MTG and for non-linguistic sounds in the angular gyrus (AG) extending into the intra-parietal sulcus (IPS). Conversely, no cortical region showed effects for congruent prime-target combinations. A subsequent effective connectivity analysis suggested that incongruency effects were likely mediated by bottom-up processes selective for each target type progressing from early auditory regions to the pSTS/MTG and AG/IPS, respectively.

Due to the nature of their design, Noppeney et al. (2007) tapped into cortical processes related to semantic priming, which might consist in a decoding of semantic content in stimuli presented in different sensory modalities on a higher, possibly cog-

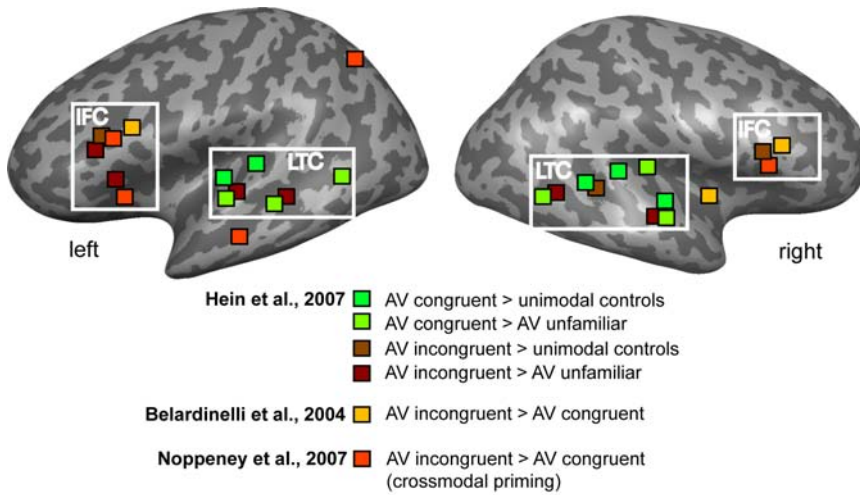


Figure 2. Regions showing significant effects for congruent and incongruent AV stimuli of common objects. Activation peaks were taken from fMRI studies investigating the effects of semantic congruency (greenish colors) and incongruency (reddish colors) of AV stimuli of common objects. These peak activations are shown on cortex reconstructions of both cerebral hemispheres of the MNI template brain. A color code for the different studies is provided in the figure. Please note that we selected only those studies which provided the coordinates of the respective activation peaks. Only activations on the lateral cortical surface are shown. See Table 1 for additional information. Abbreviations: inferior frontal cortex (IFC) and lateral temporal cortex (LTC).

nitive level of processing. However, this might have precluded an investigation of perceptual AV integration effects which could be restricted to simultaneous stimulation via both sensory channels (see van Atteveldt et al., 2007a). In a recent fMRI study by our group (Hein et al., 2007), we intended to examine both the effects of AV integration and the influence of semantic relations between the two stimulus components. For the former, we presented stimuli in auditory, visual, and AV conditions, and for the latter we used familiar, semantically related AV animal stimuli on the one hand and artificial, unfamiliar AV stimuli (lacking any strong semantic relation) on the other hand. Contrasting effects of AV stimulation with those for each unimodal condition (separately for familiar and unfamiliar stimuli), we were able to demonstrate AV integration for incongruent familiar stimuli in IFC and pSTS regions and for congruent familiar stimuli in pSTS and superior temporal gyrus (STG). In contrast, unfamiliar AV stimuli lacking any congruency relation induced activation in right IFC only. Together these findings suggest that congruency in familiar, possibly overlearned AV stimuli is processed primarily in a network comprising lateral temporal and inferior frontal regions. Most likely, the main components of the network are tuned differently to congruent and incongruent AV stimulation, with a higher sensitivity of the STG to semantically matching combinations (see also van Atteveldt et al., 2004) and of the IFC for mismatching combinations (see also Noppeney et al., 2007). However, as mentioned before, especially with respect to findings related to incongruent stimulation future studies should certainly further explore the degree of specificity of the respective effects to the crossmodal domain.

Finally, additional evidence for a functional distinction within the fronto-temporal network has been provided by work on the plasticity of AV associations. An earlier study (Gonzalo et al., 2000) found activation in the left precuneus and the right dorsolateral prefrontal cortex for learned (“congruent”) AV associations compared to non-learned (“incongruent”) combinations, while the reverse contrast revealed a widespread cortical network including anterior cingulate regions. The latter is in line with findings from our group (Naumer et al., 2008) demonstrating that even a short period (~30 minutes) of crossmodal association learning of artificial, but still object-like AV stimuli suffices to produce responses in inferior frontal regions upon the presentation of incongruent (non-learned) AV combinations. Conversely, activation shifts in the STS were found during an AV learning task (but also in a visual-to-visual learning control) as learning proceeded (Tanabe et al., 2005), suggesting that this region is critically involved during the formation of new AV associations. However, no direct comparisons between concurrent AV stimuli and unimodal controls were possible in the study of Tanabe et al. (2005), since stimuli were presented in a delayed match-to-sample task. Thus, it remains to be determined if prolonged learning of novel AV associations causes the occurrence of congruency effects for matching crossmodal stimuli in a similar manner as previously reported for overlearned material (see e.g. van Atteveldt et al., 2004; Hein et al., 2007).

Taken together, fMRI studies using AV object stimuli have identified a cortical network of temporal and frontal regions presumably mediating the respective AV integration effects. To a first approximation, some data suggest a further differentiation of this network based on a preference for either congruent or incongruent AV stimulation. As reviewed in the following section, electrophysiological studies support this differentiation at least to some degree.

### **Electrophysiological studies**

A growing body of recent electrophysiological studies has reported effects of cross-modal integration on different cortical levels. Intracranial recordings in non-human primates (e.g. Ghazanfar et al., 2005; Lakatos et al., 2007) as well as in humans (e.g. Reale et al., 2007) have started to unravel potential mechanisms operating in superior temporal regions previously regarded as strictly auditory. The great advantage of non-invasive electrophysiological techniques such as EEG and MEG, their high temporal resolution in the millisecond range, has been utilized in a number of studies to further investigate AV integration and its semantical modulation in human subjects.

Earlier studies on the neuronal time course after concurrent AV (compared to unisensory) stimulation with abstract objects and tones have found modulations of event-related potentials (ERP) associated with processing in auditory and visual cortex (Giard and Peronnet, 1999). More recently, this work has been extended to the domain of complex natural stimuli. Similar to these earlier reports, modulations of the visual N1 component were also found for AV stimuli of common objects (Molholm et al., 2004; Yuval-Greenberg and Deouell, 2007). Only for target stimuli, congruency effects around 150 ms (visual N1) and 280 ms (“selection negativity”) were reported by Molholm et al. (2004). Furthermore, subsequent dipole modeling suggested the right occipito-temporal cortex as a possible source of these effects, which is in line with reports on visual processing of stimuli associated with a memory trace generated by previous multisensory

sory experiences (Murray et al., 2004). However, since the congruency between the auditory and visual part of the stimulus had no effect on the visual N1 component in the case of non-target stimuli, these findings by Molholm and colleagues (2004) might not necessarily reflect strong semantic associations within the AV compound stimulus.

Some sensitivity to meaning-related factors at early processing stages is supported to some extent by other electrophysiological studies. For instance, Senkowski et al. (2007a) compared effects of natural and abstract AV motion stimuli. Only the former showed early multisensory integration effects after about 120 ms over the posterior scalp, involving a network of occipital, temporal and frontal regions. In contrast, the earliest cortical effects of the abstract motion stimuli were observed after 210 ms and included inferior parietal regions.

Furthermore, new analytical tools focusing on oscillatory activity move beyond the classical event-related potential approach and have demonstrated an impact of semantic congruency during AV object processing. Generally, oscillatory activity can occur either time-locked or with a latency jitter between trials. The former, the so-called “evoked” responses, are preserved after averaging across trials. In contrast, the latter, often dubbed “induced” responses would be cancelled out in conventional analyses, but can be preserved when spectral analysis is performed on a single-trial basis (Tallon-Baudry and Bertrand, 1999). In the multisensory domain, the comparison of crossmodal and unimodal stimulation is additionally exacerbated by the problem of common activity in several conditions requiring further precautions (see Senkowski et al., 2007b). Given these difficulties in analyzing electrophysiological data in the time-frequency domain, it is not surprising that only very few studies have employed AV object stimuli. A notable exception is the recent EEG study by Yuval-Greenberg and Deouell (2007). Semantic congruency between the auditory and visual stimulus components specifically affected induced responses in the gamma band between 30-70 Hz. However, these effects were found later than those reported in the ERP studies reported above, more precisely in a timewindow between 230 and 290 ms. Interestingly, a similar time window has been reported for repetition priming on induced gamma-band activity both for visually presented familiar objects (Gruber and Müller, 2005) and words (Fiebach et al., 2005). Furthermore, these findings correlated particularly with regions of the occipito-temporal cortex and have been interpreted as reflecting the formation of stable object representations both in the unimodal visual (Gruber and Müller, 2005) as well as in the crossmodal domain (Yuval-Greenberg and Deouell, 2007). Indeed, it has been suggested that induced gamma-band responses form a correlate of activated cortical object representations (Tallon-Baudry and Bertrand, 1999), but more recently it has also been associated with cognitive processes related to attention and memory (Herrmann et al., 2004; Jensen et al., 2007). Particularly from this perspective, it remains to be determined which aspects of AV object processing might actually be reflected by oscillatory activity.

Finally, with respect to AV incongruency effects for object stimuli more negative ERP responses have reported to incongruent as compared to congruent AV stimuli at a latency of ~390 ms (e.g. Molholm et al., 2004), which is in line with the N400 component previously described in linguistic contexts (Kutas and Hillyard, 1980). More precisely, this component has also been implicated in semantic mismatch processing of object stimuli (Kutas and Federmeier, 2000). Furthermore, processing of environmental



sounds after priming with a visually presented sound and vice versa both produced N400 effects when comparing unrelated to related prime-target combinations (Orgs et al., 2006). A likely source of the N400 is located in the temporal cortex, although there is at least some fMRI evidence suggesting a contribution of inferior frontal regions (van Petten and Luca, 2006). Interestingly, the N400 component has also been reported during Stroop-like interference processes in unimodal experimental settings (Hanslmayr, 2008), yielding stronger effects for incongruent compared to congruent conditions in the anterior cingulate and, at later latencies, also the inferior frontal cortex. Thus, frontal regions typically involved in conflict monitoring and cognitive control mechanisms seem to contribute critically to the occurrence of this component.

### **Preliminary conclusion**

Neuroimaging studies employing meaningful common object stimuli report at least to some extent differential processing of congruent compared to incongruent AV stimulus combinations. Particularly results from fMRI studies reviewed in this section suggest a stronger sensitivity for incongruent AV stimulation in inferior frontal and (less consistently) parietal cortices, but also to some extent in temporal regions (Figure 2). At least participation of frontal regions during incongruent AV stimulation has also been suggested by some EEG studies. In contrast, congruent AV stimuli preferentially activated lateral (Hein et al., 2007) and possibly also ventral temporal (Belardinelli et al., 2004) regions. Furthermore, electrophysiological studies seem to corroborate this finding of congruency effects in posterior temporal regions both on the level of event-related responses as well as with regard to oscillatory activity.

In the following section we will examine if these findings are consistent with results from studies using other types of AV stimuli. However, before we do so, we would like to close this section with some precautionary notes that seem to be in place at this point. An important question concerns the degree of specificity of the reported (in-) congruency effects to the crossmodal domain. Both congruency and incongruency effects might be present in a similar manner in within-modality contexts. For instance, it could be argued that the concurrent presentation of two semantically matching unimodal stimuli might induce similarly enhanced activations in the temporal cortex. However, it is difficult to test these issues since the simultaneous presentation of two auditory stimuli potentially reduces their recognizability, while the concurrent presentation of visual stimuli very often involves a spatial mismatch in the display. Thus, possibly concurrent AV stimuli activate parallel processing streams in a manner that might not be reproducible similarly in unimodal situations. However, further research using modified approaches might become increasingly important especially with regard to the construction of appropriate unimodal control conditions (see Taylor et al., 2006, Hocking and Price, 2008 for recent suggestions) allowing to test even more strictly for crossmodal integration effects in human subjects.

Furthermore, potential task influences also have to be taken into account (see Table 1). In the reported neuroimaging studies subjects had to perform diverse behavioral tasks ranging from pure fixation tasks (Hein et al., 2007; van Atteveldt et al., 2004) to explicit semantic decisions on the presented stimuli (Hocking and Price, 2008; Noppeney et al., 2007; Taylor et al., 2006). While the former might be suitable to study processes of automatic co-activation of cortical object representations generated from

Table 1. Studies modulating semantic congruency between auditory and visual object stimuli

Study	Stimulus categories	Stimulation	Task	Congruency variation and cortical region
Laurienti et al (2003)	Various; combining animate and inanimate	Images and sounds; AV only; no baseline	Decide for every visual stimulus if subject encounters it on a weekly basis	INCON: lingual and fusiform gyrus, cuneus, inferior and middle occipital regions; CON: medial PFC, anterior cingulate
Beauchamp et al. (2004)	Exp.1: tools, animals, scrambled tools and scrambled animals Exp.2: animals, tools Exp.3: tools Exp.4: tools	Exp.1: black and white images, ripple sounds; A or V Exp.2: line drawings, sounds; A, V, AV Exp.3: video clips; A, V, AV Exp.4: like in exp.3	Exp.1: one-back same/different Exp.2.: Semantic decision (A,V) con/incon decision (AV) Exp.3: object identification (delayed responses after stimulation) Exp.4: con/incon decision	Exp. 1: No Exp. 2: Yes, effects not reported Exp.3: No Exp.4: Yes, reported only for STS: trend towards con
Belardinelli et al. (2004)	Animals, tools, "human acts"	Color images, sounds; A, V, AV	Recognition (in silence, no button press)	CON: left parahippocampal gyrus, bilateral lingual gyrus; left inferior occipital gyrus INCON: bilateral IFS, bilateral insula, right lingual gyrus
Taylor et al. (2006)	Living things, non-living things	Color images, sounds, spoken words; A, V, AV	Con/incon decision	AV INCON: trend in perirhinal cortex
Naghavi et al. (2007)	Living things, non-living things	Line-drawings, sounds	Fixation	CON: right claustrum/insula
Noppeney et al. (2007)	Animals, tools	Priming experiment; words, color images (primes); spoken words sounds (targets)	Semantic decision (Is the target stimulus heavier than 4 kg?)	two types of incon (identity and response); INCON: anterior cingulate/medial PFC, bilateral IFS, left insula, IPS/AG, MTG/STS; in mPFC and left IFS stronger incon effects for both sounds and spoken words; in left STS/MTG stronger incon for words compared to sounds and the reverse effect for IPS/AG
Hein et al. (2007)	Animals, abstract objects	A,V, AV (familiar, unfamiliar)	Fixation	CON: bilateral STG, STS/MTG; INCON: largely overlapping temporal regions, bilateral IFC

Overview of recent fMRI studies using crossmodal stimuli of common objects with additional variations in semantic congruency. Abbreviations: A auditory, V visual, AV audio-visual, CON congruency, INCON incongruency, PFC prefrontal cortex, STG superior temporal gyrus, STS superior temporal sulcus, MTG middle temporal gyrus, IFC inferior frontal cortex, IFS inferior frontal sulcus, IPS intra-parietal sulcus, AG angular gyrus.



different sensory modalities, the latter operate on a higher, rather cognitive level. Thus, task effects as well as consequences of the chosen experimental design (blocked vs. event-related; see van Atteveldt et al., 2007b) have to be taken into account for the interpretation of the reported findings. We will return to these issues in the final section of this paper.

## **Cortical networks for AV integration of language-related objects**

Associations on the level of content between complex auditory and visual stimuli can also be found in domains more closely related to language. Language itself could be regarded as a system of symbols which transmits meaning independently from the sensory modality used for communication. This does not rule out that some cortical regions that are involved in language processing might fulfill a similar task in the domain of AV integration, for instance left IFC (Hagoort, 2005). However, in this section we rather concentrate on those language-related processes that are coupled more closely to the perceptual level such as AV speech and voice/face associations. Since for both domains excellent reviews have been published more recently (see Campbell, 2008; Campanella and Belin, 2007), we will not cover these topics in length but rather concentrate on semantic modulations of AV stimuli and their cortical correlates.

### **Audio-visual speech**

On the behavioral level Sumbly and Pollack (1954) reported a facilitation of speech comprehension under adverse perceptual circumstances when subjects received the combined auditory (spoken message) and visual (talker's face) input. These behavioral effects are possibly mediated by several factors since many elements of spoken language are associated with corresponding muscular facial configurations. These associations could be effective on a neuronal level down to cortical regions of the auditory cortex (Calvert et al., 1997; Pekkola et al., 2005) and are possibly modulated by attention to the visually presented articulations (Pekkola et al., 2006b). Interestingly, multi-sensory integration seems to take place only when auditory signals are perceived as speech (as opposed to the perception of the same stimuli as non-speech) and when they have been combined with concurrent video clips (Tuomainen et al., 2005).

Several neuroimaging studies have implicated a left-lateralized network of temporal regions comprising posterior portions of STG, STS, and MTG in both unimodal as well as AV speech processing (e.g. Wright et al., 2003; Callan et al., 2004). In this network, the left pSTS seems to be particularly sensitive to AV speech. First, this region exhibits superadditive responses in bimodal compared to unimodal experimental conditions and second, these effects are only present in the case of congruent AV stimulation (Calvert et al., 2000). In contrast, incongruent AV stimulation causes a drop of the fMRI signal in pSTS below the sum of the signals induced by the two unimodal conditions, i.e. a "crossmodal signal depression". These effects are in line with single-cell recordings from the dorsal and ventral banks of STS in macaques (Barraclough et al., 2005) in which 23 percent of neurons responded more strongly to congruent AV presentation of action-related stimuli. Taken together, these results from AV speech further support

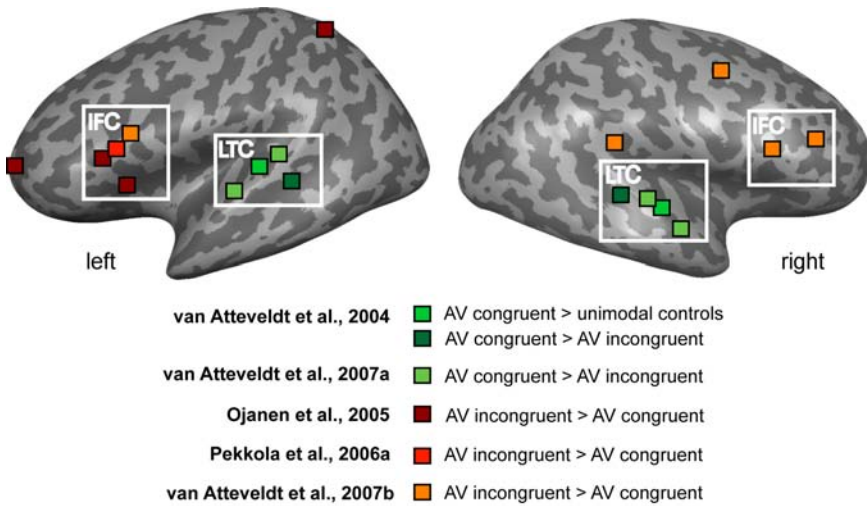


Figure 3. Regions showing significant effects for congruent and incongruent language-related AV stimuli. The same cortex representations as in Figure 2 are used for the projection of effects. Activation peaks in fMRI experiments using semantically congruent (greenish colors) and incongruent stimuli (red-dish colors) are shown only for those studies reporting coordinates. A color code for the different studies is provided in the figure. Only activations on the lateral cortical surface are shown. Abbreviations: inferior frontal cortex (IFC) and lateral temporal cortex (LTC).

the stronger responsiveness of temporal regions to congruent AV stimulation which we have suggested from our review of studies which employed common objects.

However, while the effects of congruency in these studies were determined by criteria derived from electrophysiological recordings (see Stein and Meredith, 1993), direct comparisons between congruent and incongruent AV speech stimuli provide additional evidence while controlling for equivalent stimulation complexity. Recently, Stekelenburg and Vroomen (2007) were able to demonstrate stronger effects of congruent AV speech stimuli on the auditory P2 ERP component which is assumed to be generated close to secondary auditory cortex in the superior temporal gyrus (Altmann et al., 2008). Conversely, Ojanen and others (2005) showed enhanced fMRI responses to the presentation of incongruent AV vowel stimuli in the superior parietal lobe as well as in regions of the anterior and inferior frontal cortex (Fig. 3). The latter IFC result has been replicated with the same paradigm in a group of fluent readers (Pekkola et al., 2006a).

Generally, these findings are consistent with results from studies using the so-called McGurk effect (McGurk and MacDonald, 1976) - possibly the most frequently cited evidence for speech-related AV integration. Briefly, it is generated by the presentation of video clips which induce illusory percepts of hearing /da/ although subjects actually see lips mouthing /ga/ along with auditory stimuli containing a phoneme such as /ba/. In an fMRI study by Jones and Callan (2003) cortical responses to these illusory percepts were contrasted to congruent AV speech stimuli. Effects for the incongruent McGurk condition were found in posterior parietal, inferior frontal and superior temporal regions.

Similarly, using MEG Kaiser et al. (2005) demonstrated effects in the oscillatory gamma-band activity induced by McGurk-like stimuli in the posterior parietal cortex beginning 160 ms after auditory stimulus onset, followed by occipital (270 ms) and inferior frontal (320 ms) regions. This is in contrast to the time course of event-related magnetic fields (ERFs) in response to unimodal visual speech. Activation for visual speech progresses from occipital over temporal and parietal cortices to inferior frontal regions (see Nishitani and Hari, 2002; Campbell, 2008). Thus, incongruency in AV speech induced by the McGurk effect seems to be particularly pronounced in parietal and inferior frontal regions similarly to what has been reported for AV stimuli of common objects.

However, findings derived from work on the McGurk effect have to be interpreted with some caution due to several reasons. For instance, it should be noted that cortical networks revealed by incongruency in the McGurk effect might not be easily related to findings in other domains of object-related AV integration. AV incongruency during the McGurk effect is often not explicitly perceived as a mismatch between auditory and visual inputs, but rather experienced as a fused coherent percept which does not correspond to any of its unimodal components. Interestingly, Miller and D'Esposito (2005) were able to dissociate cortical networks sensitive to sensory correspondence of AV speech stimuli from those involved during the actual perceptual fusion. In the latter network, STS showed a positive response both to fused and unfused AV stimuli, while the IPS and IFC exhibited a complementary response pattern with a stronger involvement of the parietal cortex during fused and the IFC during unfused processing. However, the unfused case might be more akin to other types of AV incongruency effects reviewed before, suggesting again a central role of the IFC in this type of cortical processing.

Additionally, besides this possible participation of IFC in incongruency processing per se, recent studies suggest its implication in motor-related activation during AV speech perception. For instance, Skipper et al. (2007) found different response patterns in the frontal cortex depending on the type of perception during a McGurk-like stimulation. When an auditory percept was categorized as corresponding to the visual component of the stimulus (thus reflecting "visual capture"), activation was found in middle and inferior frontal regions as well as the insula. However, a percept-based classification of the AV speech stimulus along the lines of the McGurk effect evoked activation in spatially adjacent but distinct right inferior and superior parietal lobules, left ventral premotor, primary motor and left somatosensory cortices. Thus, the perception of McGurk-like stimuli possibly includes cortical networks also involved in speech production.

Taken together, despite the apparent variety of results in studies investigating AV speech, findings are to some extent consistent with our hypothesis of a clear functional differentiation of temporal and frontal cortical regions with the former being more involved in processing of congruent and the latter of incongruent AV stimuli.

### **Language-related AV stimuli with arbitrary associations**

As already mentioned before, speech stimuli might be a special case of AV associations because of strong spatio-temporal interrelations between their auditory and visual components, which make certain configurations of these components at least highly unlikely. In contrast, other language-related stimuli are characterized by more arbitrary

AV associations and might therefore be more akin to the previously discussed common objects. Studies on face/voice and letter/sound associations will be covered in the following sections as two important examples of more arbitrarily associated language-related stimuli.

Congruency between AV stimuli during face-voice processing could be modulated in a number of ways. From the perspective of daily life, behaviorally most important might be cases in which affective states in facial expression and voice have to be integrated (see Ethofer et al., 2006). Accordingly, shorter reaction times were measured at least to some extent in studies in which congruent emotional information was presented in both sensory modalities (de Gelder and Vroomen, 2000). In line with these findings, the N1 ERP component, which has been implicated in auditory stimulus processing around 100 ms, was similarly enhanced in this kind of congruent AV stimulation (Pourtois et al., 2000). Effects of incongruent emotional contents in faces and voices, on the other hand, have been reported as a more pronounced mismatch negativity around 180 ms (de Gelder et al., 1999).

Neuroimaging studies measuring hemodynamic signals have often revealed modulations by congruent AV stimulation with emotional face/voice stimuli in brain regions generally implicated in processing emotional stimuli, such as the amygdala (Dolan et al., 2001). These responses in cortical and subcortical structures to emotional AV stimuli might therefore not reflect perceptual processing during crossmodal integration, but rather effects of this special type of stimulus content. However, similar to the ERP results reported above, a region close to auditory cortex in bilateral STG has recently been found to be activated more strongly by emotional stimuli than by neutral controls and, in addition, showed enhanced responses to AV stimuli compared to unimodal stimulation (Kreifelts et al., 2007). Although distinctive acoustical features of these stimuli have certainly contributed to these effects (Wiethoff et al., 2008), early AV processing in superior temporal regions seems to be modulated by the specific emotional content of the stimuli. Evidence for congruency effects for emotional face/voice stimuli, on the other hand, is rather sparse. Ethofer et al. (2006) reported stronger activation of the right posterior insula/claustum for congruent AV stimulation when compared both to incongruent and unimodal experimental conditions (see also Naghavi et al., 2007). However, the exact functional role of these regions in the course of AV processing remains rather elusive.

The second important domain of face/voice processing is related to the fact that both face and voice stimuli communicate important information with respect to speaker identity. In unimodal contexts, face processing has been linked to the fusiform face area (FFA; Kanwisher et al., 1997) in the ventral temporal cortex, while voice processing has been associated with a region in the anterior STS (Belin et al., 2000), sometimes called the temporal voice area (TVA; von Kriegstein and Giraud, 2006). It has been suggested (von Kriegstein et al., 2005) that recognition of a familiar speaker's voice critically involves a direct functional coupling between the visual FFA and the auditory TVA. However, recent neuroimaging data from a prosopagnosic patient (von Kriegstein et al., 2006) demonstrated an intact FFA-TVA interaction despite a clear behavioral deficit in the recognition of familiar speakers' voices thus suggesting that such coupling is not a sufficient neural correlate for this capability. Furthermore, a functional coupling between these two cortical areas and, additionally, between the TVA

and the IFC emerged after subjects learned face-voice (but not face-name) associations (von Kriegstein and Giraud, 2006). Interestingly, while a subsequent behavioral facilitation for the recognition of those voices involved during this learning procedure was present afterwards, no such effects were detected for the recognition of auditory control stimuli (ring tones) with learned associations to corresponding videos (of cellular phones). Thus, the findings by von Kriegstein and Giraud might be specific for the domain of face-voice processing, for which distinctive unisensory cortical correlates such as FFA and TVA are present in human subjects.

The latter studies have provided some evidence that a crossmodal context, for instance during learning, could critically modulate subsequent processing of unimodal stimuli. The recognition of an individual's identity from an auditory as well as from a visual stimulus is in any case a fascinating crossmodal capacity of human subjects which involves processing of semantic information. However, as also noted by Campanella and Belin (2007) future studies in this domain of multisensory research will certainly have to provide additional evidence from concurrent face/voice stimulation (including some congruency modulation) to investigate integrative effects more directly (see e.g. Schweinberger et al., 2007). This work would offer valuable insights, for instance with regard to the question which regions are involved in AV object processing in general or show at least some degree of specialization for certain stimulus categories.

Finally, in contrast to naturally occurring face-voice correspondences, letters and speech sounds form an example for culturally invented and transmitted AV associations that are arbitrary, but highly overlearned (others might be found for larger linguistic units such as words). Consistent with its role as a generally important AV convergence region, STS has also been found to be an important site for letter/speech integration (Raij et al., 2000; Hashimoto and Sakai, 2004; van Atteveldt et al., 2004). In line with this observation, the earliest interaction effects between both modalities have been found to occur at latencies of around 225 ms (Raij et al., 2000), which is compatible with STS as the cortical source and comparable to findings on unimodally presented visual and auditory word stimuli (Marinkovic et al., 2003).

Responsiveness for semantic congruency has been found at later delays between 380 and 450 ms after stimulus presentation (Raij et al., 2000) with an estimated source localization corresponding again to bilateral STS. However, fMRI studies (e.g. van Atteveldt et al., 2004; 2007a) suggest a stronger activation for semantically congruent (compared to incongruent) letter/speech stimuli in STG, the bilateral planum temporale and Heschl's sulcus (Figure 3). Possibly, these effects are mediated by an automatic co-activation of cortical representations for both (visual) letter and (auditory) speech stimuli (Froyen et al., 2008). Interestingly, the congruency effect could be diminished or even be completely absent when a task during the fMRI session required an active crossmodal matching of stimuli (van Atteveldt et al., 2007b). Conversely, in the same study incongruency effects have been found in an inferior fronto-parietal network suggesting at least some degree of task-dependency for these effects to occur.

### **Preliminary Conclusion**

The reviewed language-related literature shows a substantial degree of consistency with what we have discussed before with regard to AV processing of common objects (Fig. 2). Most importantly, effects of congruency and incongruency were particularly

found in lateral temporal and inferior frontal cortices (Fig. 3), respectively. However, these effects again seem to reflect functionally different processes. In superior temporal regions semantically congruent stimuli produce a stronger response while the inferior frontal cortex is more involved during the processing of incongruent AV combinations. We suggest that from a functional perspective, effects in temporal regions for congruent AV stimuli might reflect an activation of highly overlearned associations between the sight and the sound of an object, that is, a stable AV object representation. Although there are reports of single multisensory neurons in ventral prefrontal cortex of non-human primates (Sugihara et al., 2006), the reported stronger responsiveness of similar regions in human subjects could be caused by a variety of possible perceptual and cognitive processes. In the final section we thus aim to further extend our tentative explanatory account by combining the reviewed findings on AV integration with the prevailing model of semantic memory.

### **Semantic memory and AV object processing**

It is beyond the scope of an article on object-related AV integration to present or develop a complete account of semantic memory, which is one of the most intriguing cognitive abilities. Therefore, we will confine ourselves to present a rather schematic framework which will then be combined with findings on multisensory integration as reviewed in the preceding sections.

As already briefly mentioned in the introduction, semantic memory refers to stored knowledge about objects and their properties, word meanings and facts in general, which can be rapidly retrieved and forms, thereby, the basis for conceptual processing during thought and language production (cf. Martin and Chao, 2001). Essentially, knowledge about objects incorporates to large degrees knowledge about their respective sensory properties both in the unisensory and the multisensory domain. It is generally accepted that object knowledge is stored in a widely distributed network of cortical regions in the lateral and ventral temporal lobe (Martin, 2007) and, for action knowledge, additionally in fronto-parietal regions. Its distributed nature notwithstanding, this network is organized in a systematic manner, for instance according to categorical distinctions between living and non-living objects which seems to hold true both for the visual (Chao et al., 1999) and the auditory domains (Lewis et al., 2005; Murray et al., 2006). Importantly, these cortical structures are activated not only upon stimulus presentation, but also under purely conceptual demands, for example during property verification tasks (Chao et al., 1999). Furthermore, there is evidence that these cognitive processes operate in a sensory-specific manner, critically involving the respective cortical regions which are activated by sensory stimulation (Goldberg et al., 2006).

Multiple instances of concurrent stimulus presentations, during which, for example, visual and acoustical features are experienced simultaneously, form the critical learning history resulting in an increase in associative strength between cortical representations of these unimodal components. Furthermore, it appears to be likely that the resulting AV object representation is stored in regions of the temporal lobe. Given the strong association between the unimodal subcomponents (at least for some types of material such as auditory and visual speech stimuli; Calvert et al., 1997) the presenta-



tion of one component in isolation might suffice for the co-activation of the cortical representation of its counterparts in other sensory modalities. In any case, combined congruent AV stimulation seems to activate an object representation in a more pronounced manner, thereby causing additional activation even down to cortical regions traditionally regarded as unisensory (see e.g. van Atteveldt et al. (2004); Hein et al. (2007), Ghazanfar et al. (2005)). Furthermore, it has been repeatedly shown that temporal regions such as the left pMTG are also involved in the storage and representation of action knowledge (see e.g. Kellenbach et al., 2003; Lewis, 2006). In line with the notion that the combined representation of auditory and visual stimuli might be further enhanced by the additional association of related actions (like for tool stimuli), both temporal and frontal regions seem to be modulated by the congruency of AV action stimuli (Galati et al., 2008).

In sum, the complex interplay of (multi-)sensory information encoding during semantic memory formation and the retrieval of object information upon later presentations of AV stimuli, which is additionally guided by the respective task, provides a challenging background for research to disentangle these different processes. Based on the literature reviewed above, a reasonable position would be that enhanced activation due to semantically congruent AV stimulation reflects a more pronounced activation of the respective cortical object representation than in unimodal or incongruent cases. This hypothesis still remains preliminary since future studies combining different experimental techniques and paradigms using various tasks are needed to further distinguish multisensory (perceptual) stimulus processing from other, rather amodal (cognitive) processes (see e.g. Hocking and Price, 2008 for recent suggestions in this context).

While these proposed mechanisms might account for congruency effects found in temporal regions, another rather consistent finding from the studies reviewed here was the stronger IFC activation for semantically incongruent AV stimuli. It has been suggested previously (Martin and Chao, 2001; Wagner et al., 2001) that the IFC plays an important role for semantic retrieval and working memory processes. Studies using fMRI (e.g. Wagner et al., 2000) and transcranial magnetic stimulation (TMS; e.g. Gough et al., 2005) particularly implicated the anterior portion of the IFG for semantic processing. Interestingly, it has been argued (Badre et al., 2005; Badre and Wagner, 2007) that the anterior IFG controls access to stored conceptual representations, especially in cases where the presented items are only weakly associated. Thus, increased control demands in these situations might be reflected in stronger anterior IFG activations, while rather automatic processes (with reduced IFG activations) could operate in the case of strongly associated material. Conflicting crossmodal stimuli might similarly induce increased control demands, thereby inducing stronger IFG activations.

However, it should be noted that most multisensory experiments are not optimized to distinguish between this explanation for an IFC involvement and possible alternative accounts. For instance, it has been argued that inferior frontal regions might be more responsive to abstract compared to concrete semantic demands (Goldberg et al., 2007), to selection processes (Thompson-Schill, 2003), to conflicting information (Egner and Hirsch, 2005; Noppeney et al., 2007) or to object-based categorization (Jiang et al., 2007). Additionally, an involvement of frontal regions conjointly with regions of the parietal cortex has been demonstrated in crossmodal contexts whenever se-

lective or divided attention has to be deployed (see e.g. Weissman et al., 2004; Johnson and Zatorre, 2006). Thus, future studies in the field of multisensory research focusing on semantic aspects of crossmodally presented stimuli will certainly benefit from an exchange with scientists working on theoretical accounts of frontal cortex function. Furthermore, it has been convincingly argued for additional cortical regions to be implicated in semantic processing, particularly the anterior temporal lobe (Patterson et al., 2007; Pobric et al., 2007). We expect that an improved differentiation between sensory processing (both within and across modalities) and higher-order cognitive processing based on the “pure” content of stimuli irrespective of which sensory modality has provided this content, will lead to significant further progress at the intersection of perception and cognition.

## **Overall conclusion**

We have reviewed the available neuroimaging literature investigating the impact of meaning (“semantics”) on processes of audio-visual integration both on the behavioral level and with respect to its cortical correlates. A number of studies have revealed cortical regions of the temporal and frontal lobes which are implicated 1) in AV integration and 2) in the processing of congruent and incongruent content. We suggest that temporal regions are more involved for matching AV stimulation possibly due to an enhanced activation of stable AV object representations. In contrast, inferior frontal regions participate more in the processing of incongruent stimuli, potentially reflecting increased cognitive control demands. Future studies will have to investigate if the fronto-temporal gradient we propose truly reflects an organizational principle by determining, for instance, if it remains stable across different types of tasks and by further elucidating the relative contributions of bottom-up and top-down processes. Also, given the ubiquitous presence of lateral frontal activations in conflict tasks employing one single stimulus modality only, it may be conclusive to examine whether temporal and frontal activations can be dissociated on grounds of their specificity to crossmodal stimulation.

## **Acknowledgements**

Work related to this paper was supported by the intramural young investigator program of Frankfurt Medical School (M.J.N.). The authors would like to thank Christoph Bledowski, Jochen Kaiser, Benjamin Rahm, Sarah Weigelt and three anonymous reviewers for their helpful comments on earlier versions of the paper.





# S

## Summary

Stimulus-related processing in the brain can be characterized by the two major operations analysis and integration. "Analysis" refers to those processes in which complex stimuli are segregated into more basic features, while "integration" could be regarded as the generation of cortical representations with sufficient degrees of complexity to enable cognition and action. However, in the brain these operations interact in a highly dynamic manner and involve widely distributed networks of subcortical and cortical regions. The present thesis focuses particularly on those integrative processes in the human cerebral cortex operating on complex auditory and audio-visual stimuli. Furthermore, the experimental work described here had the additional aim to investigate the contribution of semantics to sensory integration. Thus, the overall set of questions which inspired the work collected in the present thesis can be stated as follows: First, how is sensory integration accomplished by the human brain? Secondly, what is the specific contribution of stimulus content (meaning) for this kind of processing? And thirdly, which cortical regions are most relevant for the generation of meaning in the sensory brain?

**Chapter 1** further describes and discusses these general ideas which were the driving forces behind the research collected in the present thesis. The aim of this introductory chapter is to offer a guide for readers who are less familiar with the respective empirical and theoretical background.

The next two chapters report two functional magnetic resonance imaging (fMRI) adaptation studies which examined the question of cortical representations for complex sounds and different auditory object categories. More specifically, in **chapter 2** cortical representations of animal vocalizations were investigated as one highly familiar and meaningful category of complex sounds. We hypothesized to find effects particularly in the superior temporal gyrus - anterior to primary auditory cortex - which contains neurons tuned to a variety of different frequencies and spectro-temporal features of sounds. Adaptation indeed occurred in the superior temporal gyrus of the left cortical hemisphere suggesting that cortical representations of animal vocalizations were driven particularly by spectro-temporal dynamics of sound stimuli. Overall, this first auditory fMRI-adaptation study demonstrates that the utilization of complex and meaningful sounds provides a key to explore the complete auditory processing hierarchy.

In the next experimental step reported in **chapter 3** we aimed at further examining the response profiles of higher-order auditory cortex for complex, meaningful sound stimuli. We took advantage of the fact that human subjects learn during the first years of their cognitive development to classify sounds as being members of different categories. Interestingly and in line with our findings, previous research by others has suggested that the anterior superior temporal gyrus is particularly involved in the processing of animal vocalizations while a cortical region located on the left posterior middle temporal gyrus responds stronger to sounds of man-made objects such as tools. Again employing an fMRI-adaptation paradigm, we replicated our results from the first study with respect to an adaptation effect for animal vocalizations in left superior temporal gyrus. However, no corresponding adaptation effect was found in this cortical region for repeated versus different tool sounds, thereby supporting the view of selectivity for animal vocalizations. In contrast, the left posterior middle temporal gyrus was the only cortical region which adapted selectively to repeated tool sounds. Furthermore,

these two regions seemed to operate on different levels within the processing hierarchy: The superior temporal gyrus, but not the posterior middle temporal gyrus showed responses significantly different from baseline. Thus, only the former region seems to be sensitive to spectro-temporal features present in any sound of sufficient complexity which underscores the view that this cortical region indeed mediates the differentiation of sounds into distinct categories on the basis of their complex physical properties. This might then provide an input into higher-level regions which are involved in the further integration with conceptual representations.

In a second set of empirical studies we extended the scope from the processing of meaningful auditory stimuli to the question of how semantics exerts influences on the integration across sensory modalities. In **chapter 4** I present a study from our group in which we combined fMRI with a simple associative learning paradigm. In previous work we were able to demonstrate that factors such as the familiarity and the degree of semantic congruency between auditory and visual stimuli were reflected in differential cortical response patterns. We reasoned that cortical representations for initially non-meaningful, artificial audio-visual stimuli might converge to some extent after learning onto those cortical regions commonly involved in the integration of familiar audio-visual stimuli, such as animal vocalizations and corresponding images. Importantly, we hypothesized that after training these specific artificial AV combinations might be processed as being congruent and other combinations as being incongruent. In a short behavioral training session subjects learned to associate specific artificial sounds and objects. Training-induced cortical changes for the processing of artificial audio-visual stimuli were determined by fMRI before (PRE) and after (POST) training. In the PRE-session unfamiliar artificial audio-visual stimuli were particularly integrated in the right inferior frontal cortex. In the POST-training session, a comparable analysis demonstrated increased responses in bilateral inferior frontal, intra-parietal and lateral temporal regions extending into the superior temporal sulcus. A similar set of regions was detected in previous studies for the integration of familiar (animal) vocalizations and images. Furthermore, the cortical response patterns indeed differed in a systematic manner between learned and unlearned combinations of artificial audio-visual stimuli. While mismatching compared to matching (learned) audio-visual combinations induced stronger responses in left inferior frontal regions, the reverse contrast was found to be significant in posterior portions of the right middle temporal gyrus and, to a lesser extent, in right fusiform gyrus. Most interestingly, these patterns also converged onto findings for semantically incongruent compared to congruent familiar audio-visual stimuli and vice versa. Therefore our findings suggested that even after short-term training of associations between artificial sounds and images processes of integration and match- versus mismatch-detection operated in a way which is remarkably similar to the case of familiar audio-visual stimuli.

**Chapter 5** reports a study in which we extended the application of fMRI adaptation to the domain of multisensory integration. We employed four experimental conditions in which two successive audio-visual animal stimuli were presented involving either a change in the auditory, visual, both or none of the components from the first to the second stimulus. Importantly, changes occurred within the same basic-level category of animals, for instance by two different images depicting a dog, in order to reduce subtle congruency or incongruency effects within or across the two audio-visual

stimuli. In line with results from unimodal adaptation studies, significantly suppressed responses for auditory repetitions compared to auditory changes were detected in the right superior temporal gyrus and sulcus and for repeated versus changed visual stimuli in right lateral occipital regions and the fusiform gyrus. Surprisingly, higher-order sensory cortices were not insensitive to repetitions versus changes in the respective other sensory modality. More precisely, effects of auditory repetition suppression were accompanied in right superior temporal regions by stronger fMRI signals for visual repetitions than for visual changes (repetition enhancement), while a region in the left lateral occipital cortex showed enhanced responses for auditory repetitions. This finding that stimulus repetitions versus changes in one sensory modality induced differential processing in cortices of the respective other sensory modality underscored that fMRI adaptation is a highly useful tool also for research in the multisensory domain. Future studies will therefore certainly utilize this experimental technique for more advanced questions involving variations of the semantic content in the context of sensory integration.

The application of novel experimental techniques needs to be constrained by theoretical considerations in order to generate new insights in integrative cortical mechanisms. The final **chapter 6** of the present dissertation offers a first theoretical framework concerning the impact of semantics on processes of audio-visual integration. This framework is guided by observations such as those reported in chapter 4. More precisely, a fronto-temporal “gradient” has repeatedly been reported with more pronounced responses for semantically incongruent audio-visual stimulation in inferior frontal regions and stronger effects for semantically congruent audio-visual stimuli in lateral temporal cortex. Potential perceptual and cognitive processes which might underlie these differential cortical patterns are discussed in this chapter. We are confident that our framework offers a starting point for future collaborative work to further unravel the integrative mechanisms operating in the multisensory brain.

# S

## **Samenvatting**

Stimulus-gerelateerde verwerking in het brein kan worden gekarakteriseerd door twee belangrijke operaties: analyse en integratie. Analyse betreft zich op processen waarin complexe stimuli gescheiden worden in meer basale kenmerken, terwijl integratie beschouwd kan worden als de vorming van corticale representaties met voldoende complexiteit om cognitie en actie mogelijk te maken. In het brein interacteren deze operaties op zeer dynamische wijze waarbij breed gedistribueerde netwerken van subcorticale en corticale gebieden betrokken zijn. Het huidige proefschrift concentreert zich specifiek op de integratieprocessen in de menselijke hersenschors die opereren op complexe auditieve en audio-visuele stimuli. Daarbij had het experimentele werk dat hier beschreven wordt het bijkomende doel om de bijdrage van inhoudelijke betekenis aan sensorische integratie te onderzoeken. De overkoepelende vragen die het werk dat verzameld is in dit proefschrift inspireerden, kunnen als volgt geformuleerd worden: ten eerste, hoe wordt sensorische integratie bereikt door het menselijke brein? Ten tweede, wat is de specifieke bijdrage van stimulusinhoud (betekenis) aan dit soort verwerking? En ten derde, welke corticale gebieden zijn het meest relevant voor het voortbrengen van betekenis in het sensorische brein?

**Hoofdstuk 1** beschrijft en bediscussieert deze algemene ideeën die de drijvende kracht waren achter het onderzoek verzameld in dit proefschrift in meer detail. Het doel van dit inleidende hoofdstuk is om een handreiking te bieden aan de lezers die minder bekend zijn met de respectievelijke empirische en theoretische achtergrond.

De volgende twee hoofdstukken beschrijven twee fMRI (functionele magnetische resonantie tomografie) adaptatiestudies die het thema corticale representaties voor complexe geluiden en verschillende auditieve objectcategorieën onderzochten. Om precies te zijn werden in **hoofdstuk 2** corticale representaties onderzocht voor dieren-geluiden, die een welbekende en betekenisvolle categorie van complexe geluiden vormen. We hadden de hypothese dat we vooral effecten zouden vinden in de superieure temporale gyrus – anterior gelegen ten opzichte van de primaire auditieve cortex – die neuronen bevat die reageren op een verscheidenheid van verschillende frequenties en spectro-temporele kenmerken van geluiden. Adaptatie trad inderdaad op in de superieure temporale gyrus van de linker hersenhelft, wat erop duidt dat corticale representaties van dieregeluiden specifiek gedreven worden door de spectro-temporele dynamiek van geluiden. Meer algemeen laat deze eerste auditieve fMRI-adaptatiestudie zien dat het gebruik van complexe en betekenisvolle auditieve stimuli een ingang biedt om de volledige auditieve verwerkingshiërarchie te bestuderen.

De volgende experimentele stap, gerapporteerd in **hoofdstuk 3** was erop gericht om de responseprofielen van de hogere-orde auditieve cortex voor complexe, betekenisvolle geluiden verder te bestuderen. Daarbij maakten we gebruik van het feit dat mensen in de eerste jaren van hun cognitieve ontwikkeling leren om geluiden in verschillende categorieën in te delen. Het is interessant en in overeenstemming met onze bevindingen dat eerder onderzoek door anderen aangeeft dat de anterior superieure temporale cortex specifiek betrokken is bij de verwerking van dieregeluiden terwijl een corticaal gebied in de linker posterieure middelste temporele gyrus sterker op geluiden van kunstmatige objecten zoals gereedschap reageert. Opnieuw gebruikmakend van een fMRI-adaptatieparadigma repliceerden we de resultaten van onze eerste studie wat betreft het adaptatie-effect voor dieregeluiden in de linker superieure temporale gyrus. Desondanks vonden we geen overeenkomstig adaptatie-

effect voor herhaalde versus verschillende gereedschapsgeluiden in dit corticale gebied, wat de visie van selectiviteit voor dierengeluiden ondersteunt. Daarentegen was de linker posterieure middelste temporale gyrus het enige corticale gebied dat selectief adapteerde aan herhaalde geluiden van gereedschap. Bovendien leken deze beide gebieden op verschillende niveaus in de verwerkingshiërarchie te functioneren: de superieure temporale gyrus, maar niet de posterieure middelste temporale gyrus liet responsen zien die significant van de baseline verschilden. Dus alleen het eerstgenoemde gebied lijkt gevoelig te zijn voor de spectro-temporele kenmerken die elk geluid van voldoende complexiteit bezit, wat de visie versterkt dat dit corticale gebied inderdaad betrokken is bij de differentiatie van geluiden in verschillende categorieën op basis van hun complexe fysieke eigenschappen. Dit zou dan een input aan hogere-orde gebieden verschaffen die betrokken zijn bij de verdere integratie met conceptuele representaties.

In een tweede serie empirische studies breidden we ons onderzoeksterrein uit van de verwerking van betekenisvolle auditieve stimuli tot de kwestie van hoe betekenis invloed uitoefent op de integratie over sensorische modaliteiten. In **hoofdstuk 4** presenteer ik een studie van onze groep waarin we fMRI combineerden met een simpel associatief-leren paradigma. In eerder werk konden we laten zien dat factoren als de bekendheid van en de mate van semantische overeenstemming tussen auditieve en visuele stimuli weerspiegeld werden in differentiële corticale activatiepatronen. We redeneerden dat corticale representaties voor in eerste instantie betekenisloze en kunstmatige audio-visuele (AV) stimuli na leren tot op zekere hoogte zouden kunnen convergeren in die corticale gebieden die normaalgesproken betrokken zijn bij de integratie van bekende audio-visuele stimuli, zoals dierengeluiden en bijbehorende plaatjes. We hadden de hypothese dat na training deze specifieke kunstmatige AV combinaties verwerkt zouden kunnen worden als congruent en andere combinaties als incongruent. In een korte gedragsmatige trainingssessie leerden proefpersonen associaties tussen specifieke artificiële geluiden en objecten. Training-geïnduceerde corticale veranderingen voor de verwerking van artificiële audio-visuele stimuli werden bepaald met fMRI vóór (PRE) en ná (POST) training. In de PRE-training sessie werden onbekende kunstmatige audio-visuele stimuli voornamelijk geïntegreerd in de rechter inferieure frontale cortex. In de POST-training sessie, liet een vergelijkbare analyse verhoogde activatie zien in bilaterale inferieure frontale, intra-parietale en lateraal- temporale gebieden zich uitstrekkend tot in de superieure temporale sulcus. Een vergelijkbare groep gebieden werd in eerdere studies voor de integratie van bekende geluiden en beelden van dieren en mensen gevonden. Verder verschilden de corticale activatiepatronen inderdaad op een systematische manier voor geleerde en niet-geleerde combinaties van kunstmatige audio-visuele stimuli. Terwijl niet-overeenstemmende vergeleken met overeenstemmende (geleerde) audio-visuele combinaties sterkere responsen op wekten in de linker inferieure frontale gebieden, vonden we dat het omgekeerde contrast significant was in posterieure delen van de rechter middelste temporale gyrus en, in mindere mate, in de rechter fusiforme gyrus. Het meest interessant was dat deze patronen overeenkwamen met bevindingen voor semantisch incongruente tegenover congruente bekende audio-visuele stimuli en vice versa. Daarom geven onze bevindingen aan dat zelfs na korte-termijn training van associaties tussen kunstmatige geluiden en beelden integratieprocessen en detectie



van overeenstemming op opmerkelijk vergelijkbare wijze functioneren als in het geval van bekende audio-visuele stimuli.

**Hoofdstuk 5** bericht over een studie waarin we de toepassing van fMRI adaptatie uitbreidden naar het domein van multisensorische integratie. We gebruikten vier experimentele condities waarin twee opeenvolgende audio-visuele dierenstimuli aangeboden werden met een verandering in ofwel de auditieve, visuele, beide, of geen enkele component van de eerste ten opzichte van de tweede stimulus. Het is belangrijk op te merken dat de veranderingen optraden in dezelfde basale categorie van dieren, bijvoorbeeld twee verschillende afbeeldingen van een hond, om subtiele congruentie- of incongruentie-effecten binnen of tussen de twee audio-visuele stimuli te verminderen. In overeenstemming met resultaten van unimodale adaptatiestudies werden significant onderdrukte responsen gedetecteerd in de rechter superieure temporale gyrus en sulcus voor auditieve herhalingen vergeleken met auditieve veranderingen en in de rechter laterale occipitale gebieden en de fusiforme gyrus voor herhaalde versus veranderde visuele stimuli. Verrassend genoeg waren hogere-orde sensorische cortices niet ongevoelig voor herhalingen versus veranderingen in de respectievelijke andere modaliteit. Om precies te zijn gingen effecten van auditieve onderdrukking in de rechter superieure temporale cortex samen met sterkere fMRI signalen voor visuele herhalingen dan voor visuele veranderingen (repetitie-versterking), terwijl een gebied in de linker laterale occipitale cortex versterkte responsen vertoonde voor auditieve herhalingen. Deze bevinding dat stimulusherhalingen versus -veranderingen in één sensorische modaliteit leiden tot differentiële verwerking in cortices van de respectievelijke andere sensorische modaliteit onderstreepte dat fMRI-adaptatie ook een zeer nuttig middel is voor multisensorisch onderzoek. Toekomstige studies zullen daarom zeker van deze experimentele techniek voor meer geavanceerde vragen met betrekking tot variaties in betekenisinhoud in de context van sensorische integratie gebruikmaken.

De toepassing van nieuwe experimentele technieken moet gemotiveerd zijn door theoretische overwegingen om nieuwe inzichten in corticale integratiemechanismen te kunnen voortbrengen. Het laatste **hoofdstuk 6** van dit proefschrift biedt een eerste theoretisch kader wat betreft de impact van betekenis op processen van audio-visuele integratie. Dit kader wordt opgebouwd aan de hand van observaties zoals gerapporteerd in hoofdstuk 4. Om precies te zijn is er herhaaldelijk een fronto-temporale "gradient" gerapporteerd met meer uitgesproken responsen voor semantisch incongruente audio-visuele stimulatie in inferieure frontale gebieden en sterkere effecten voor semantisch congruente audio-visuele stimuli in laterale temporale cortex. Mogelijke perceptuele en cognitieve processen die aan deze differentiële corticale patronen ten grondslag zouden kunnen liggen worden in dit hoofdstuk bediscussieerd. We zijn vol vertrouwen dat ons kader een startpunt biedt voor toekomstig gemeenschappelijk werk om de integratiemechanismen van het multisensorische brein te ontrafelen.

# R

## References

- Adams RB, Janata P (2002) A comparison of neural circuits underlying auditory and visual object categorization. *NeuroImage* 16:361-377.
- Ahveninen J, Jaaskelainen IP, Raji T, Bonmassar G, Devore S, Hamalainen M, Levanen S, Lin FH, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW (2006) Task-modulated “what” and “where” pathways in human auditory cortex. *Proc Natl Acad Sci U S A* 103:14608-14613.
- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL (2001) “What” and “where” in the human auditory system. *Proc Natl Acad Sci U S A* 98:12301-12306.
- Alsius A, Navarra J, Campbell R, Soto-Faraco S (2005) Audiovisual integration of speech falters under high attention demands. *Curr Biol* 15:839-843.
- Altmann CF, Deubelius A, Kourtzi Z (2004) Shape saliency modulates contextual processing in the human lateral occipital complex. *J Cogn Neurosci* 16:794-804.
- Altmann CF, Doehrmann O, Kaiser J (2007) Selectivity for animal vocalizations in the human auditory cortex. *Cereb Cortex* 17:2601-2608.
- Altmann CF, Nakata H, Noguchi Y, Inui K, Hoshiyama M, Kaneoke Y, Kakigi R (2008) Temporal dynamics of adaptation to natural sounds in the human auditory cortex. *Cereb Cortex* 18:1350-1360.
- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559-571.
- Amedi A, Stern WM, Camprodon JA, Bermpohl F, Merabet L, Rotman S, Hemond C, Meijer P, Pascual-Leone A (2007) Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat Neurosci* 10:687-689.
- Arnott SR, Binns MA, Grady CL, Alain C (2004) Assessing the auditory dual-pathway model in humans. *NeuroImage* 22:401-408.
- Badgaiyan RD, Schacter DL, Alpert NM (1999) Auditory priming within and across modalities: evidence from positron emission tomography. *J Cogn Neurosci* 11:337-348.
- Badgaiyan RD, Schacter DL, Alpert NM (2001) Priming within and across modalities: exploring the nature of rCBF increases and decreases. *NeuroImage* 13:272-282.
- Badre D, Wagner AD (2007) Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45:2883-2901.
- Badre D, Poldrack RA, Pare-Blagoev EJ, Insler RZ, Wagner AD (2005) Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47:907-918.
- Baehrick LE (1992) Infants’ perceptual differentiation of amodal and modality-specific audio-visual relations. *Journal of Experimental Child Psychology* 53:180-199.
- Baehrick LE (1994) The development of infants’ sensitivity to arbitrary intermodal relations. *Ecological Psychology* 6:111-123.
- Barracough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci* 17:377-391.
- Beauchamp MS (2005) Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93-114.
- Beauchamp MS, Lee KE, Haxby JV, Martin A (2002) Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34:149-159.

- Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809-823.
- Belardinelli MO, Sestieri C, Di Matteo R, Delogu F, Del Gratta C, Ferretti A, Caulo M, Tartaro A, Romani GL (2004) Audio-visual crossmodal interactions in environmental perception: an fMRI investigation. *Cognitive Processing* 5:167-174.
- Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14:2105-2109.
- Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. *Brain Res Cogn Brain Res* 13:17-26.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309-312.
- Belin P, Zilbovicius M, Crozier S, Thivard L, Fontaine A, Masure MC, Samson Y (1998) Lateralization of speech and auditory temporal processing. *J Cogn Neurosci* 10:536-540.
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Stat Soc B* 57: 289-300.
- Bergerbest D, Ghahremani DG, Gabrieli JD (2004) Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *J Cogn Neurosci* 16:966-977.
- Binder JR, Liebenthal E, Possing ET, Medler DA, Ward BD (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci* 7:295-301.
- Binder JR, Frost JA, Hammeke TA, Bellgowan PSF, Springer JA, Kaufman JN, Possing ET (2000) Human temporal lobe activation by speech and nonspeech sounds. *Cereb Cortex* 10:512-528.
- Boemio A, Fromm S, Braun A, Poeppel D (2005) Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nat Neurosci* 8:389-395.
- Botvinick MM, Cohen JD, Carter CS (2004) Conflict monitoring and anterior cingulate cortex: an update. *Trends Cog Sci* 8:539-546.
- Boynton GM, Finney EM (2003) Orientation-specific adaptation in human visual cortex. *J Neurosci* 23:8781-8787.
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207-4221.
- Brechmann A, Baumgart F, Scheich H (2002) Sound-level-dependent representation of frequency modulations in human auditory cortex: a low-noise fMRI study. *J Neurophysiol* 87:423-433.
- Brugge JF, Merzenich MM (1973) Responses of neurons in auditory cortex of the macaque monkey to monaural and binaural stimulation. *J Neurophysiol* 36:1138-1158.
- Brugge JF, Volkov IO, Garell PC, Reale RA, Howard MA, III (2003) Functional connections between auditory cortex on Heschl's gyrus and on the lateral superior temporal gyrus in humans. *J Neurophysiol* 90:3750-3763.
- Buckner RL, Koutstaal W, Schacter DL, Rosen BR (2000) Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain* 123:620-640.

- Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004) Multi-sensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J Cogn Neurosci* 16:805-816.
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110-1123.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649-657.
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *NeuroImage* 14:427-438.
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593-596.
- Campanella S, Belin P (2007) Integrating face and voice in person perception. *Trends Cogn Sci* 11:535-543.
- Campbell R (2008) The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:1001-1010.
- Carter CS, van Veen V (2007) Anterior cingulate cortex and conflict detection: an update of theory and data. *Cogn Affect Behav Neurosci* 7:367-379.
- Chao LL, Haxby JV, Martin A (1999) Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2:913-919.
- Christoffels IK, Formisano E, Schiller NO (2007) Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum Brain Mapp* 28:868-879.
- Clarke S, Bellmann A, Meuli RA, Assal G, Steck AJ (2000) Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia* 38:797-807.
- Clarke S, Bellmann Thiran A, Maeder P, Adriani M, Vernet O, Regli L, Cuisenaire O, Thiran JP (2002) What and where in human audition: selective deficits following focal hemispheric lesions. *Exp Brain Res* 147:8-15.
- Clavagnier S, Falchier A, Kennedy H (2004) Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cogn Affect Behav Neurosci* 4:117-126.
- Clemo RH, Sharma G, Allman B, Meredith AM (2008) Auditory projections to extrastriate visual cortex: connectional basis for multisensory processing in 'unimodal' visual neurons. *Exp Brain Res* 191:37-47.
- Colavita F (1974) Human sensory dominance. *Percept Psychophys* 25:409-412.
- Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201-215.
- D'Esposito M (2007) From cognitive to neural models of working memory. *Philos Trans R Soc Lond B Biol Sci* 362:761-772.
- Dale AM (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* 8:109-114.

- Damasio H, Grabowski TJ, Tranel D, Ponto LLB, Hichwa RD, Damasio AR (2001) Neural correlates of naming actions and of naming spatial relations. *NeuroImage* 13:1053-1064.
- De Gelder B, Vroomen J (2000) The perception of emotions by ear and by eye. *Cognition and Emotion* 14:289-311.
- De Gelder B, Bocker KB, Tuomainen J, Hensen M, Vroomen J (1999) The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neurosci Lett* 260:133-136.
- Degerman A, Rinne T, Pekkola J, Autti T, Jaaskelainen IP, Sams M, Alho K (2007) Human brain activity associated with audiovisual perception and attention. *NeuroImage* 34:1683-1691.
- Dehaene-Lambertz G, Dehaene S, Anton JL, Campagne A, Ciuciu P, Dehaene GP, Denghien I, Jobert A, Lebihan D, Sigman M, Pallier C, Poline JB (2006) Functional segregation of cortical language areas by sentence repetition. *Hum Brain Mapp* 27:360-371.
- Desimone R (1996) Neural mechanisms for visual memory and their role in attention. *Proc Natl Acad Sci U S A* 93:13494-13499.
- Doehrmann O, Naumer MJ (2008) Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Res.*
- Doehrmann O, Naumer MJ, Volz S, Kaiser J, Altmann CF (2008) Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia* 46:2776-2786.
- Dolan RJ, Morris JS, de Gelder B (2001) Crossmodal binding of fear in voice and face. *Proc Natl Acad Sci U S A* 98:10006-10010.
- Downing PE, Chan AW, Peelen MV, Dodds CM, Kanwisher N (2006) Domain specificity in visual cortex. *Cereb Cortex* 16:1453-1461.
- Driver J, Noesselt T (2008) Multisensory interplay reveals crossmodal influences on 'sensory-specific' brain regions, neural responses, and judgments. *Neuron* 57:11-23.
- Duncan J, Owen AM (2000) Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci* 23:475-483.
- Egner T, Hirsch J (2005) Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat Neurosci* 8:1784-1790.
- Engel AK, Fries P, Konig P, Brecht M, Singer W (1999) Temporal binding, binocular rivalry, and consciousness. *Conscious Cogn* 8:128-151.
- Engelien A, Sibersweig D, Stern E, Huber W, Doring W, Frith C, Frackowiak RSJ (1995) The functional anatomy of recovery from auditory agnosia: A PET study of sound categorization in a neurological patient and normal controls. *Brain* 118:1395-1409.
- Engelien A, Tuscher O, Hermans W, Isenberg N, Eidelberg D, Frith C, Stern E, Silbersweig D (2006) Functional neuroanatomy of non-verbal semantic sound processing in humans. *J Neural Transm* 113:599-608.
- Ernst MO, Bulthoff HH (2004) Merging the senses into a robust percept. *Trends Cogn Sci* 8:162-169.
- Ethofer T, Anders S, Erb M, Droll C, Royen L, Saur R, Reiterer S, Grodd W, Wildgruber D (2006) Impact of voice on emotional judgment of faces: an event-related fMRI study. *Hum Brain Mapp* 27:707-714.

- Falchier A, Clavagnier S, Barone P, Kennedy H (2002) Anatomical evidence of multimodal integration in primate striate cortex. *J Neurosci* 22:5749-5759.
- Fecteau S, Armony JL, Joanette Y, Belin P (2004) Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage* 23:840-848.
- Fecteau S, Armony JL, Joanette Y, Belin P (2005) Sensitivity to voice in human prefrontal cortex. *J Neurophysiol* 94:2251-2254.
- Fiebach CJ, Gruber T, Supp GG (2005) Neuronal mechanisms of repetition priming in occipitotemporal cortex: spatiotemporal evidence from functional magnetic resonance imaging and electroencephalography. *J Neurosci* 25:3414-3422.
- Fischl B, Sereno MI, Tootell RB, Dale AM (1999) High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272-284.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 33:636-647.
- Formisano E, Kim DS, Di Salle F, van de Moortele PF, Ugurbil K, Goebel R (2003) Mirror-symmetric tonotopic maps in human primary auditory cortex. *Neuron* 40:859-869.
- Forster B, Cavina-Pratesi C, Aglioti SM, Berlucchi G (2002) Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Exp Brain Res* 143:480-487.
- Frassinetti F, Bolognini N, Ladavas E (2002) Enhancement of visual perception by cross-modal visuo-auditory interaction. *Exp Brain Res* 147:332-343.
- Friston KJ, Josephs O, Rees G, Turner R (1998) Nonlinear event-related responses in fMRI. *Magn Reson Med* 39:41-52.
- Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ (1995a) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 2:189-210.
- Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R (1995b) Analysis of fMRI time-series revisited. *NeuroImage* 2:45-53.
- Froyen D, Van Attevelde N, Bonte M, Blomert L (2008) Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neurosci Lett* 430:23-28.
- Fujisaki W, Shimojo S, Kashino M, Nishida S (2004) Recalibration of audiovisual simultaneity. *Nat Neurosci* 7:773-778.
- Fuster JM, Jervey JP (1981) Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science* 212:952-955.
- Fuster JM, Bodner M, Kroger JK (2000) Cross-modal and cross-temporal association in neurons of frontal cortex. *Nature* 405:347-351.
- Gaab N, Gabrieli JD, Glover GH (2007a) Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. *Hum Brain Mapp* 28:703-720.
- Gaab N, Gabrieli JD, Glover GH (2007b) Assessing the influence of scanner background noise on auditory processing. II. An fMRI study comparing auditory processing in the absence and presence of recorded scanner noise using a sparse design. *Hum Brain Mapp* 28:721-732.



- Gagnepain P, Chetelat G, Landeau B, Dayan J, Eustache F, Lebreton K (2008) Spoken word memory traces within the human auditory cortex revealed by repetition priming and functional magnetic resonance imaging. *J Neurosci* 28:5281-5289.
- Galati G, Committeri G, Spitoni G, Aprile T, Di Russo F, Pitzalis S, Pizzamiglio L (2008) A selective representation of the meaning of actions in the auditory mirror system. *NeuroImage* 40:1274-1286.
- Gallace A, Spence C (2006) Multisensory synesthetic interactions in the speeded classification of visual size. *Perception & Psychophysics* 68:1191-1203.
- Ghazanfar AA, Schroeder CE (2006) Is neocortex essentially multisensory? *Trends Cogn Sci* 10:278-285.
- Ghazanfar AA, Chandrasekaran C, Logothetis NK (2008) Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J Neurosci* 28:4457-4469.
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004-5012.
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473-490.
- Gibson JR, Maunsell JH (1997) Sensory modality specificity of neural activity related to memory in visual cortex. *J Neurophysiol* 78:1263-1275.
- Giraud AL, Price CJ (2001) The constraints functional neuroimaging places on classical models of auditory word processing. *J Cogn Neurosci* 13:754-765.
- Giraud AL, Lorenzi C, Ashburner J, Wable J, Johnsrude I, Frackowiak R, Kleinschmidt A (2000) Representation of the temporal envelope of sounds in the human brain. *J Neurophysiol* 84:1588-1598.
- Goebel R, Esposito F, Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with Brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Hum Brain Mapp* 27:392-401.
- Gogate LJ, Bahrick LE (1998) Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *J Exp Child Psychol* 69:133-149.
- Goldberg RF, Perfetti CA, Schneider W (2006) Perceptual knowledge retrieval activates sensory brain regions. *J Neurosci* 26:4917-4921.
- Goldberg RF, Perfetti CA, Fiez JA, Schneider W (2007) Selective retrieval of abstract semantic knowledge in left prefrontal cortex. *J Neurosci* 27:3790-3798.
- Gonzalo D, Shallice T, Dolan R (2000) Time-dependent changes in learning audiovisual associations: a single-trial fMRI study. *NeuroImage* 11:243-255.
- Gottfried JA, Dolan RJ (2003) The nose smells what the eye sees: crossmodal visual facilitation of human olfactory perception. *Neuron* 39:375-386.
- Gough PM, Nobre AC, Devlin JT (2005) Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *J Neurosci* 25:8010-8016.
- Griffiths TD, Warren JD, Scott SK, Nelken I, King AJ (2004) Cortical processing of complex sound: a way forward? *Trends Neurosci* 27:181-185.



- Grill-Spector K, Malach R (2001) fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107:293-321.
- Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10:14-23.
- Grill-Spector K, Kushnir T, Edelman S, Avidan G, Itzchak Y, Malach R (1999) Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24:187-203.
- Gruber T, Muller MM (2002) Effects of picture repetition on induced gamma band responses, evoked potentials, and phase synchrony in the human EEG. *Brain Res Cogn Brain Res* 13:377-392.
- Gruber T, Muller MM (2005) Oscillatory brain activity dissociates between associative stimulus content in a repetition priming task in the human EEG. *Cereb Cortex* 15:109-116.
- Hackett TA, Stepniewska I, Kaas JH (1998) Subdivisions of auditory cortex and ipsilateral cortical connections of the parabelt auditory cortex in macaque monkeys. *J Comp Neurol* 394:475-495.
- Hagoort P (2005) On Broca, brain, and binding: a new framework. *Trends Cogn Sci* 9:416-423.
- Hall DA, Johnsrude IS, Haggard MP, Palmer AR, Akeroyd MA, Summerfield AQ (2002) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 12:140-149.
- Hall DA, Haggard MP, Akeroyd MA, Palmer AR, Summerfield AQ, Elliott MR, Gurney EM, Bowtell RW (1999) "Sparse" temporal sampling in auditory fMRI. *Hum Brain Mapp* 7:213-223.
- Hanslmayr S, Pastotter B, Bauml K-H, Gruber S, Wimber M, Klimesch W (2008) The electrophysiological dynamics of interference during the stroop task. *J Cogn Neurosci* 20:215-225.
- Hart HC, Palmer AR, Hall DA (2003) Amplitude and frequency-modulated stimuli activate common regions of human auditory cortex. *Cereb Cortex* 13:773-781.
- Hashimoto R, Sakai KL (2004) Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron* 42:311-322.
- Hein G, Doehrmann O, Muller NG, Kaiser J, Muckli L, Naumer MJ (2007) Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J Neurosci* 27:7881-7887.
- Henson R, Shallice T, Dolan R (2000) Neuroimaging evidence for dissociable forms of repetition priming. *Science* 287:1269-1272.
- Henson RN (2003) Neuroimaging studies of priming. *Prog Neurobiol* 70:53-81.
- Herrmann CS, Knight RT (2001) Mechanisms of human attention: event-related potentials and oscillations. *Neurosci Biobehav Rev* 25:465-476.
- Herrmann CS, Munk MH, Engel AK (2004) Cognitive functions of gamma-band activity: memory match and utilization. *Trends Cogn Sci* 8:347-355.
- Hocking J, Price CJ (2008) The role of the posterior superior temporal sulcus in audiovisual processing. *Cereb Cortex* 18:2439-2449.

- Howard MA, 3rd, Volkov IO, Abbas PJ, Damasio H, Ollendieck MC, Granner MA (1996) A chronic microelectrode investigation of the tonotopic organization of human auditory cortex. *Brain Res* 724:260-264.
- Huettel SA, McCarthy G (2000) Evidence for a refractory period in the hemodynamic response to visual stimuli as measured by MRI. *NeuroImage* 11:547-553.
- Inui K, Okamoto H, Miki K, Gunji A, Kakigi R (2006) Serial and parallel processing in the human auditory cortex: a magnetoencephalographic study. *Cereb Cortex* 16:18-30.
- Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levanen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW (2004) Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci U S A* 101:6809-6814.
- James TW, Gauthier I (2003) Auditory and action semantic features activate sensory-specific perceptual brain regions. *Curr Biol* 13:1792-1796.
- James TW, Humphrey GK, Gati JS, Menon RS, Goodale MA (2002a) Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron* 35:793.
- James TW, Humphrey GK, Gati JS, Servos P, Menon RS, Goodale MA (2002b) Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia* 40:1706-1714.
- Jamison HL, Watkins KE, Bishop DVM, Matthews PM (2006) Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb Cortex* 16:1266-1275.
- Jancke L, Wustenberg T, Scheich H, Heinze HJ (2002) Phonetic perception and the temporal cortex. *NeuroImage* 15:733-746.
- Jensen O, Kaiser J, Lachaux JP (2007) Human gamma-frequency oscillations associated with attention and memory. *Trends Neurosci* 30:317-324.
- Jiang X, Bradley E, Rini RA, Zeffiro T, VanMeter J, Riesenhuber M (2007) Categorization training results in shape- and category-selective human neural plasticity. *Neuron* 53:891-903.
- Johnson-Frey SH, Newman-Norlund R, Grafton ST (2005) A distributed left hemisphere network active during planning of everyday tool use skills. *Cereb Cortex* 15:681-695.
- Johnson JA, Zatorre RJ (2006) Neural substrates for dividing and focusing attention between simultaneous auditory and visual events. *NeuroImage* 31:1673-1681.
- Jones EG, Dell'Anna ME, Molinari M, Rausell E, Hashikawa T (1995) Subdivisions of macaque monkey auditory cortex revealed by calcium-binding protein immunoreactivity. *J Comp Neurol* 362:153-170.
- Jones JA, Callan DE (2003) Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. *Neuroreport* 14:1129-1133.
- Kaas JH, Hackett TA (2000) Subdivisions of auditory cortex and processing streams in primates. *Proc Natl Acad Sci U S A* 97:11793-11799.
- Kaas JH, Hackett TA, Tramo MJ (1999) Auditory processing in primate cerebral cortex. *Curr Opin Neurobiol* 9:164-170.
- Kaiser J, Lutzenberger W (2005) Cortical oscillatory activity and the dynamics of auditory memory processing. *Rev Neurosci* 16:239-254.

- Kaiser J, Lutzenberger W, Ackermann H, Birbaumer N (2002) Dynamics of gamma-band activity induced by auditory pattern changes in humans. *Cereb Cortex* 12:212-221.
- Kaiser J, Ripper B, Birbaumer N, Lutzenberger W (2003) Dynamics of gamma-band activity in human magnetoencephalogram during auditory pattern working memory. *NeuroImage* 20:816-827.
- Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W (2005) Hearing lips: gamma-band activity during audiovisual speech perception. *Cereb Cortex* 15:646-653.
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302-4311.
- Kayser C, Logothetis NK (2007) Do early sensory cortices integrate cross-modal information? *Brain Struct Funct* 212:121-132.
- Kellenbach ML, Brett M, Patterson K (2003) Actions speak louder than functions: the importance of manipulability and action in tool representation. *J Cogn Neurosci* 15:30-46.
- Koppen C, Alsius A, Spence C (2008) Semantic congruency and the Colavita visual dominance effect. *Exp Brain Res* 184:533-546.
- Kourtzi Z, Kanwisher N (2001) Representation of perceived object shape by the human lateral occipital complex. *Science* 293:1506-1509.
- Kourtzi Z, Betts LR, Sarkheil P, Welchman AE (2005) Distributed neural plasticity for shape learning in the human visual cortex. *PLoS Biol* 3:e204.
- Kourtzi Z, Tolias AS, Altmann CF, Augath M, Logothetis NK (2003) Integration of local features into global shapes: monkey and human fMRI studies. *Neuron* 37:333-346.
- Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D (2007) Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage* 37:1445-1456.
- Krekelberg B, Boynton GM, van Wezel RJ (2006) Adaptation: from single cells to BOLD signals. *Trends Neurosci* 29:250-256.
- Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207:203-205.
- Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci* 4:463-470.
- Lakatos P, Chen C-M, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279-292.
- Laurienti PJ, Kraft RA, Maldjian JA, Burdette JH, Wallace MT (2004) Semantic congruence is a critical factor in multisensory behavioral performance. *Exp Brain Res* 158:405-414.
- Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE (2005) On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Exp Brain Res* 166:289-297.
- Laurienti PJ, Wallace MT, Maldjian JA, Susi CM, Stein BE, Burdette JH (2003) Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. *Hum Brain Mapp* 19:213-223.
- Lehmann S, Murray MM (2005) The role of multisensory memories in unisensory object discrimination. *Brain Res Cogn Brain Res* 24:326-334.
- Lerner Y, Hendler T, Ben-Bashat D, Harel M, Malach R (2001) A hierarchical axis of object processing stages in the human visual cortex. *Cereb Cortex* 11:287-297.

- Lewis JW (2006) Cortical networks related to human use of tools. *Neuroscientist* 12:211-231.
- Lewis JW, Phinney RE, Brefczynski-Lewis JA, DeYoe EA (2006) Lefties get it “right” when hearing tool sounds. *J Cogn Neurosci* 18:1314-1330.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005) Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* 25:5148-5158.
- Lewis JW, Wightman FL, Brefczynski JA, Phinney RE, Binder JR, DeYoe EA (2004) Human brain regions involved in recognizing environmental sounds. *Cereb Cortex* 14:1008-1021.
- Li L, Miller EK, Desimone R (1993) The representation of stimulus familiarity in anterior inferior temporal cortex. *J Neurophysiol* 69:1918-1929.
- Lickliter R, Bahrick LE (2004) Perceptual development and the origins of multisensory responsiveness. In: *Handbook of Multisensory Processes* (Calvert GA, Spence C, Stein BE, eds), pp 643-654. Cambridge, MA: MIT Press.
- Liegeois-Chauvel C, Lorenzi C, Trebuchon A, Regis J, Chauvel P (2004) Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* 14:731-740.
- Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S (2001) Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14:802-816.
- Mahon BZ, Milleville SC, Negri GAL, Rumiati RI, Caramazza A, Martin A (2007) Action-related properties shape object representations in the ventral stream. *Neuron* 55:507.
- Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proc Natl Acad Sci U S A* 92:8135-8139.
- Marcell MM, Borella D, Greene M, Kerr E, Rogers S (2000) Confrontation naming of environmental sounds. *Journal of Clinical Experimental Neuropsychology* 22:830-864.
- Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E (2003) Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* 38:487-497.
- Martin A (2007) The representation of object concepts in the brain. *Ann Rev Psychol* 58:25-45.
- Martin A, Chao LL (2001) Semantic memory and the brain: structure and processes. *Curr Opin Neurobiol* 11:194-201.
- Martin A, Wiggs CL, Ungerleider LG, Haxby JV (1996) Neural correlates of category-specific knowledge. *Nature* 379:649-652.
- McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746-748.
- Meienbrock A, Naumer MJ, Doehrmann O, Singer W, Muckli L (2007) Retinotopic effects during spatial audio-visual integration. *Neuropsychologia* 45:531-539.
- Mesulam MM (1998) From sensation to cognition. *Brain* 121 (Pt 6):1013-1052.
- Miller EK, Li L, Desimone R (1991) A neural mechanism for working and recognition memory in inferior temporal cortex. *Science* 254:1377-1379.
- Miller J (1982) Divided attention: evidence for coactivation with redundant signals. *Cognit Psychol* 14:247-279.

- Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci* 25:5884-5893.
- Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb Cortex* 14:452-465.
- Morel A, Garraghty PE, Kaas JH (1993) Tonotopic organization, architectonic fields, and connections of auditory cortex in macaque monkeys. *J Comp Neurol* 335:437-459.
- Morrongiello BA, Fenwick KD, Chance G (1998) Crossmodal learning in newborn infants: inferences about properties of auditory-visual events. *Infant Behavior and Development* 21:543-553.
- Möttönen R, Calvert GA, Jaaskelainen IP, Matthews PM, Thesen T, Tuomainen J, Sams M (2006) Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *NeuroImage* 30:563-569.
- Murray MM, Camen C, Spierer L, Clarke S (2008) Plasticity in representations of environmental sounds revealed by electrical neuroimaging. *NeuroImage* 39:847-856.
- Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S (2006a) Rapid brain discrimination of sounds of objects. *J Neurosci* 26:1293-1302.
- Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, Schnider A (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *NeuroImage* 21:125-135.
- Murray SO, Wojciulik E (2004) Attention increases neural selectivity in the human lateral occipital complex. *Nat Neurosci* 7:70-74.
- Murray SO, Olman CA, Kersten D (2006b) Spatially specific fMRI repetition effects in human visual cortex. *J Neurophysiol* 95:2439-2445.
- Naghavi HR, Eriksson J, Larsson A, Nyberg L (2007) The claustrum/insula region integrates conceptually related sounds and pictures. *Neurosci Lett* 422:77-80.
- Naumer MJ, Doehrmann, O, Müller, NG, Muckli, L, Kaiser, J, Hein, G (2008) Cortical plasticity of audio-visual object representations. *Cereb Cortex*: 10.1093/cercor/bhn200.
- Navarra J, Soto-Faraco S, Spence C (2007) Adaptation to audiotactile asynchrony. *Neuroscience Letters* 413:72.
- Nelken I (2004) Processing of complex stimuli and natural scenes in the auditory cortex. *Curr Opin Neurobiol* 14:474-480.
- New J, Cosmides L, Tooby J (2007) Category-specific attention for animals reflects ancestral priorities, not expertise. *Proc Natl Acad Sci U S A* 104:16598-16603.
- Nishitani N, Hari R (2002) Viewing lip forms: cortical dynamics. *Neuron* 36:1211-1220.
- Noppeney U, Price CJ, Penny WD, Friston KJ (2006) Two distinct neural mechanisms for category-selective responses. *Cereb Cortex* 16:437-445.
- Noppeney U, Josephs O, Kiebel S, Friston KJ, Price CJ (2005) Action selectivity in parietal and temporal cortex. *Brain Res Cog Brain Res* 25:641-649.
- Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ (2008) The effect of prior visual information on recognition of speech and sounds. *Cereb Cortex* 18:598-609.
- Nyberg L, Sandblom J, Jones S, Neely AS, Petersson KM, Ingvar M, Backman L (2003) Neural correlates of training-related memory improvement in adulthood and aging. *Proc Natl Acad Sci U S A* 100:13728-13733.
- Ojanen V, Mottonen R, Pekkola J, Jaaskelainen IP, Joensuu R, Autti T, Sams M (2005) Processing of audiovisual speech in Broca's area. *NeuroImage* 25:333-338.

- Olesen PJ, Westerberg H, Klingberg T (2004) Increased prefrontal and parietal activity after training of working memory. *Nat Neurosci* 7:75-79.
- Op de Beeck HP, Haushofer J, Kanwisher NG (2008) Interpreting fMRI data: maps, modules and dimensions. *Nat Rev Neurosci* 9:123-135.
- Orgs G, Lange K, Dombrowski JH, Heil M (2006) Conceptual priming for environmental sounds and words: an ERP study. *Brain Cogn* 62:267-272.
- Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8:976-987.
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767-776.
- Pekkola J, Laasonen M, Ojanen V, Autti T, Jaaskelainen IP, Kujala T, Sams M (2006a) Perception of matching and conflicting audiovisual speech in dyslexic and fluent readers: An fMRI study at 3 T. *NeuroImage* 29:797-807.
- Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Sams M (2006b) Attention to visual speech gestures enhances hemodynamic activity in the left planum temporale. *Hum Brain Mapp* 27:471-477.
- Pekkola J, Ojanen V, Autti T, Jaaskelainen IP, Mottonen R, Tarkiainen A, Sams M (2005) Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport* 16:125-128.
- Petkov CI, Kayser C, Augath M, Logothetis NK (2006) Functional imaging reveals numerous fields in the monkey auditory cortex. *PLoS Biol* 4:e215.
- Petkov CI, Kang X, Alho K, Bertrand O, Yund EW, Woods DL (2004) Attentional modulation of human auditory cortex. *Nat Neurosci* 7:658-663.
- Pobric G, Jefferies E, Ralph MA (2007) Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. *Proc Natl Acad Sci U S A* 104:20137-20141.
- Pourtois G, de Gelder B, Vroomen J, Rossion B, Crommelinck M (2000) The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport* 11:1329-1333.
- Price C, Thierry G, Griffiths T (2005) Speech-specific auditory processing: where is it? *Trends Cog Sci* 9:271-276.
- Putzar L, Goerendt I, Lange K, Rösler F, Röder B (2007) Early visual deprivation impairs multisensory interactions in humans. *Nat Neurosci* 10:1243-1245.
- Rademacher J, Morosan P, Schormann T, Schleicher A, Werner C, Freund HJ, Zilles K (2001) Probabilistic mapping and volume measurement of human primary auditory cortex. *NeuroImage* 13:669-683.
- Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. *Neuron* 28:617-625.
- Rauschecker JP (1997) Processing of complex sounds in the auditory cortex of cat, monkey, and man. *Acta Otolaryngol Suppl* 532:34-38.
- Rauschecker JP (1998a) Parallel processing in the auditory cortex of primates. *Audiol Neurotol* 3:86-103.
- Rauschecker JP (1998b) Cortical processing of complex sounds. *Curr Opin Neurobiol* 8:516-521.



- Rauschecker JP, Tian B (2000) Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A* 97:11800-11806.
- Rauschecker JP, Tian B, Hauser M (1995) Processing of complex sounds in the macaque nonprimary auditory cortex. *Science* 268:111-114.
- Rauschecker JP, Tian B, Pons T, Mishkin M (1997) Serial and parallel processing in rhesus monkey auditory cortex. *J Comp Neurol* 382:89-103.
- Reale RA, Calvert GA, Thesen T, Jenison RL, Kawasaki H, Oya H, Howard MA, Brugge JF (2007) Auditory-visual processing represented in the human superior temporal gyrus. *Neuroscience* 145:162-184.
- Romanski LM, Goldman-Rakic PS (2002) An auditory domain in primate prefrontal cortex. *Nat Neurosci* 5:15-16.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP (1999) Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci* 2:1131-1136.
- Sawamura H, Orban GA, Vogels R (2006) Selectivity of neuronal adaptation does not match response selectivity: a single-cell study of the fMRI adaptation paradigm. *Neuron* 49:307-318.
- Sawamura H, Georgieva S, Vogels R, Vanduffel W, Orban GA (2005) Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. *J Neurosci* 25:4294-4306.
- Schmidt CF, Zaehle T, Meyer M, Geiser E, Boesiger P, Jancke L (2008) Silent and continuous fMRI scanning differentially modulate activation in an auditory language comprehension task. *Hum Brain Mapp* 29:46-56.
- Schonwiesner M, Rubsamen R, von Cramon DY (2005) Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur J Neurosci* 22:1521-1528.
- Schroeder CE, Foxe J (2005) Multisensory contributions to low-level, ‘unisensory’ processing. *Curr Opin Neurobiol* 15:454-458.
- Schwartz S, Maquet P, Frith C (2002) Neural correlates of perceptual learning: a functional MRI study of visual texture discrimination. *Proc Natl Acad Sci U S A* 99:17137-17142.
- Schwarzbauer C, Davis MH, Rodd JM, Johnsrude I (2006) Interleaved silent steady state (ISSS) imaging: a new sparse imaging method applied to auditory fMRI. *NeuroImage* 29:774-782.
- Schweinberger SR, Robertson D, Kaufmann JM (2007) Hearing facial identities. *Q J Exp Psychol (Colchester)* 60:1446-1456.
- Scott SK (2005) Auditory processing - Speech, space and auditory objects. *Curr Opin Neurobiol* 15:197-201.
- Seifritz E, Di Salle F, Esposito F, Herdener M, Neuhoﬀ JG, Scheﬄer K (2006) Enhancing BOLD response in the auditory system by neurophysiologically tuned fMRI sequence. *NeuroImage* 29:1013-1022.
- Seitz AR, Kim R, Shams L (2006) Sound facilitates visual learning. *Curr Biol* 16:1422-1427.
- Senkowski D, Saint-Amour D, Kelly SP, Foxe JJ (2007a) Multisensory processing of naturalistic objects in motion: A high-density electrical mapping and source estimation study. *NeuroImage* 36:877-888.

- Senkowski D, Gomez-Ramirez M, Lakatos P, Wylie GR, Molholm S, Schroeder CE, Foxe JJ (2007b) Multisensory processing and oscillatory activity: analyzing non-linear electrophysiological measures in humans and simians. *Exp Brain Res* 177:184-195.
- Sestieri C, Di Matteo R, Ferretti A, Del Gratta C, Caulo M, Tartaro A, Olivetti Belardinelli M, Romani GL (2006) "What" versus "Where" in the audiovisual domain: An fMRI study. *NeuroImage* 33:672-680.
- Shams L, Seitz AR (2008) Benefits of multisensory learning. *Trends Cogn Sci* 12:411-417.
- Shomstein S, Yantis S (2004) Control of attention shifts between vision and audition in human cortex. *J Neurosci* 24:10702-10706.
- Singer W (1999) Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24:49-65, 111-125.
- Skipper JI, van Wassenhove V, Nusbaum HC, Small SL (2007) Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex* 17:2387-2399.
- Smith EL, Grabowecky M, Suzuki S (2007) Auditory-visual crossmodal integration in perception of face gender. *Current Biology* 17:1680-1685.
- Spiers HJ, Maguire EA (2007) Decoding human brain activity during real-world experiences. *Trends Cog Sci* 11:356-365.
- Stanford TR, Stein BE (2007) Superadditivity in multisensory integration: putting the computation in context. *Neuroreport* 18:787-792.
- Stein BE, Meredith MA, eds (1993) *The merging of the senses*. Cambridge, MA: MIT Press.
- Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. *Nat Rev Neurosci* 9:255-266.
- Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. *J Cogn Neurosci* 19:1964-1973.
- Stevenson R, Geoghegan M, James T (2007) Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research* 179:85-95.
- Sugihara T, Diltz MD, Averbach BB, Romanski LM (2006) Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *J Neurosci* 26:11138-11147.
- Sugita Y, Suzuki Y (2003) Audiovisual perception: Implicit estimation of sound-arrival time. *Nature* 421:911.
- Sumby WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212-215.
- Sweet RA, Dorph-Petersen KA, Lewis DA (2005) Mapping auditory core, lateral belt, and parabelt cortices in the human superior temporal gyrus. *J Comp Neurol* 491:270-289.
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. New York: Thieme Medical Publishers.
- Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences* 3:151-162.
- Tanabe HC, Honda M, Sadato N (2005) Functionally segregated neural substrates for arbitrary audiovisual paired-association learning. *J Neurosci* 25:6409-6418.



- Taylor KI, Moss HE, Stamatakis EA, Tyler LK (2006) Binding crossmodal object features in perirhinal cortex. *Proc Natl Acad Sci U S A* 103:8239-8244.
- Thierry G, Giraud AL, Price C (2003) Hemispheric dissociation in access to the human semantic system. *Neuron* 38:499-506.
- Thompson-Schill SL (2003) Neuroimaging studies of semantic memory: inferring "how" from "where." *Neuropsychologia* 41:280-292.
- Thompson SK, von Kriegstein K, Deane-Pratt A, Marquardt T, Deichmann R, Griffiths TD, McAlpine D (2006) Representation of interaural time delay in the human auditory midbrain. *Nat Neurosci* 9:1096-1098.
- Tian B, Rauschecker JP (1998) Processing of frequency-modulated sounds in the cat's posterior auditory field. *J Neurophysiol* 79:2629-2642.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP (2001) Functional specialization in rhesus monkey auditory cortex. *Science* 292:290-293.
- Tjan BS, Lestou V, Kourtzi Z (2006) Uncertainty and invariance in the human visual cortex. *J Neurophysiol* 96:1556-1568.
- Tootell RB, Hadjikhani NK, Vanduffel W, Liu AK, Mendola JD, Sereno MI, Dale AM (1998) Functional analysis of primary visual cortex (V1) in humans. *Proc Natl Acad Sci U S A* 95:811-817.
- Tranel D, Damasio H, Damasio AR (1997) A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia* 35:1319-1327.
- Tranel D, Grabowski TJ, Lyon J, Damasio H (2005) Naming the same entities from visual or from auditory stimulation engages similar regions of left inferotemporal cortices. *J Cogn Neurosci* 17:1293-1305.
- Tranel D, Kemmerer D, Adolphs R, Damasio H, Damasio AR (2003a) Neural correlates of conceptual knowledge for actions. *Cognitive Neuropsychology* 20:409 - 432.
- Tranel D, Damasio H, Eichhorn GR, Grabowski T, Ponto LLB, Hichwa RD (2003b) Neural correlates of naming animals from their characteristic sounds. *Neuropsychologia* 41:847-854.
- Tuomainen J, Andersen TS, Tiippana K, Sams M (2005) Audio-visual speech perception is special. *Cognition* 96:813.
- Turk-Browne NB, Yi DJ, Leber AB, Chun MM (2007) Visual quality determines the direction of neural repetition effects. *Cereb Cortex* 17:425-433.
- Ulanovsky N, Las L, Nelken I (2003) Processing of low-probability sounds by cortical neurons. *Nat Neurosci* 6:391-398.
- van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43:271-282.
- van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007a) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17:962-974.
- van Atteveldt NM, Formisano E, Goebel R, Blomert L (2007b) Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36:1345-1360.
- Van Petten C, Luka BJ (2006) Neural localization of semantic context effects in electro-magnetic and hemodynamic studies. *Brain and Language* 97:279.
- Vatakis A, Spence C (2007) Crossmodal binding: Evaluating the unity assumption using audiovisual speech stimuli. *Perception & Psychophysics* 69:744.

- von Kriegstein K, Giraud A-L (2006) Implicit multisensory associations influence voice recognition. *PLoS Biology* 4:e326.
- von Kriegstein K, Kleinschmidt A, Giraud A-L (2006) Voice Recognition and Cross-Modal Responses to Familiar Speakers' Voices in Prosopagnosia. *Cereb Cortex* 16:1314-1322.
- von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud A-L (2005) Interaction of Face and Voice Areas during Speaker Recognition. *J Cogn Neurosci* 17:367-376.
- Vuilleumier P, Henson RN, Driver J, Dolan RJ (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci* 5:491-499.
- Wagner AD, Pare-Blagoev EJ, Clark J, Poldrack RA (2001) Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31:329-338.
- Wagner AD, Koutstaal W, Maril A, Schacter DL, Buckner RL (2000) Task-specific repetition priming in left inferior prefrontal cortex. *Cereb Cortex* 10:1176-1184.
- Wallace MT, Perrault TJ, Jr., Hairston WD, Stein BE (2004) Visual experience is necessary for the development of multisensory integration. *J Neurosci* 24:9580-9584.
- Wandell BA, Dumoulin SO, Brewer AA (2007) Visual field maps in human cortex. *Neuron* 56:366-383.
- Weigelt S, Singer W, Muckli L (2007a) Separate cortical stages in amodal completion revealed by functional magnetic resonance adaptation. *BMC Neuroscience* 8:70.
- Weigelt S, Muckli L, Kohler A (2008) The method of functional magnetic resonance adaptation in visual neuroscience. *Rev Neurosci*.
- Weigelt S, Kourtzi Z, Kohler A, Singer W, Muckli L (2007b) The cortical representation of objects rotating in depth. *J Neurosci* 27:3864-3874.
- Weisberg J, van Turenout M, Martin A (2007) A neural system for learning about object function. *Cereb Cortex* 17:513-521.
- Weissman DH, Warner LM, Woldorff MG (2004) The neural mechanisms for minimizing cross-modal distraction. *J Neurosci* 24:10941-10949.
- Welch RB, Warren DH (1980) Immediate perceptual response to intersensory discrepancy. *Psychol Bull* 88:638-667.
- Wessinger CM, VanMeter J, Tian B, Van Lare J, Pekar J, Rauschecker JP (2001) Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *J Cogn Neurosci* 13:1-7.
- Westbury CF, Zatorre RJ, Evans AC (1999) Quantifying Variability in the Planum Temporale: A Probability Map. *Cereb Cortex* 9:392-405.
- Whatmough C, Chertkow H, Murtha S, Hanratty K (2002) Dissociable brain regions process object meaning and object structure during picture naming. *Neuropsychologia* 40:174.
- Wheeler ME, Petersen SE, Buckner RL (2000) Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proc Natl Acad Sci U S A* 97:11125-11129.
- Wiethoff S, Wildgruber D, Kreifelts B, Becker H, Herbert C, Grodd W, Ethofer T (2008) Cerebral processing of emotional prosody-influence of acoustic parameters and arousal. *Neuroimage* 39:885-893.
- Wiggs CL, Martin A (1998) Properties and mechanisms of perceptual priming. *Curr Opin Neurobiol* 8:227-233.

## References

- Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 13:1034-1043.
- Yuval-Greenberg S, Deouell LY (2007) What You See Is Not (Always) What You Hear: Induced Gamma Band Responses Reflect Cross-Modal Interactions in Familiar Object Recognition. *J Neurosci* 27:1090-1096.
- Zatorre RJ, Belin P (2001) Spectral and temporal processing in human auditory cortex. *Cereb Cortex* 11:946-953.
- Zatorre RJ, Bouffard M, Belin P (2004) Sensitivity to auditory object features in human temporal neocortex. *J Neurosci* 24:3637-3642.

## Abbreviations

2D, 3D	two-dimensional, three-dimensional
a	anterior
A	auditory
A1	primary auditory cortex
ACC	anterior cingulate cortex
AG	angular gyrus
AM	amplitude modulation
ANOVA	analysis of variance
ART	artificial
AV	audio-visual
BOLD	blood-oxygenation level-dependent
CON	congruent
dB	decibel carrier
DLPFC	dorso-lateral prefrontal cortex
EEG	electroencephalography
EPI	echo-planar imaging
ERF	event-related field
ERP	event-related potential
FA	flip angle
FDR	false discovery rate
FFA	fusiform face area
FG	fusiform gyrus
FoV	field of view
fMRI	functional magnetic resonance imaging
fMRIa	functional magnetic resonance imaging adaptation
FWHM	full width at half maximum
GCM	Granger causality mapping
GLM	general linear model
HG	Heschl's gyrus
IFC	inferior frontal cortex
IFG	inferior frontal gyrus
IFS	inferior frontal sulcus
INCON	incongruent
IPL	inferior parietal lobule
IPS	inferior parietal sulcus
ISI	interstimulus interval
IT	inferior temporal
ITG	inferior temporal gyrus
LH	left hemisphere
LO	lateral occipital
LOC	lateral occipital complex
LTC	lateral temporal cortex
MEG	magnetoencephalography
MFC	medial frontal cortex

## References

MP-RAGE	magnetization-prepared rapid gradient-echo
MRI	magnetic resonance imaging
MTG	middle temporal gyrus
NAT	natural
p	posterior
PFC	prefrontal cortex
PMC	premotor cortex
PP	planum polare
PPC	posterior parietal cortex
PRE	before-training session
POST	after-training session
PT	planum temporale
R	tonotopic area with an organization mirror-symmetric to A1
RE	repetition enhancement
RfA	release from adaptation
RH	right hemisphere
RFX	random effects
RS	repetition suppression
ROI	region of interest
SEM	standard error of the mean
STG	superior temporal gyrus
STS	superior temporal sulcus
TE	time to echo
TR	time to repetition
TMS	transcranial magnetic stimulation
TVA	temporal voice area
V1	primary visual cortex
VO	ventral occipital